A Framework to Guide the Assessment of Human–Machine Systems

Kimberly Stowers, **James Oglesby**, University of Central Florida, Orlando, **Shirley Sonesh**, Tulane University, New Orleans, Louisiana, **Kevin Leyva**, Riptide Software, Orlando, Florida, **Chelsea Iwig**, and **Eduardo Salas**, Rice University, Houston, TX

Objective: We have developed a framework for guiding measurement in human–machine systems.

Background: The assessment of safety and performance in human–machine systems often relies on direct measurement, such as tracking reaction time and accidents. However, safety and performance emerge from the combination of several variables. The assessment of precursors to safety and performance are thus an important part of predicting and improving outcomes in human–machine systems.

Method: As part of an in-depth literature analysis involving peer-reviewed, empirical articles, we located and classified variables important to human–machine systems, giving a snapshot of the state of science on human–machine system safety and performance. Using this information, we created a framework of safety and performance in human–machine systems.

Results: This framework details several inputs and processes that collectively influence safety and performance. Inputs are divided according to human, machine, and environmental inputs. Processes are divided into attitudes, behaviors, and cognitive variables. Each class of inputs influences the processes and, subsequently, outcomes that emerge in human–machine systems.

Conclusion: This framework offers a useful starting point for understanding the current state of the science and measuring many of the complex variables relating to safety and performance in human-machine systems.

Application: This framework can be applied to the design, development, and implementation of automated machines in spaceflight, military, and health care settings. We present a hypothetical example in our write-up of how it can be used to aid in project success.

Keywords: human–machine interaction, human–robot interaction, human–computer interaction, performance, safety

---

Address correspondence to Kimberly Stowers, University of Central Florida, Orlando, FL 32826, USA; e-mail: kim.stwrs@gmail.com.

INTRODUCTION

Tim is an aerospace engineer at Ad Astra Engineering Corp (AAEC) involved in the development of robotic systems for future long-distance and long-duration exploration missions to Mars. He and his colleagues have developed an initial prototype for a new teleoperated robot, and based on his tests, the prototype seems to work perfectly. Tim has been told by his superiors that he must thoroughly evaluate the robot's safety and performance using appropriate measurement techniques. However, Tim is new at AAEC and he has never formally assessed the safety and performance of one of his machines. Not knowing where to begin, he conducts a quick online search but is overwhelmed with the results. Some researchers measure user-centered variables, such as situation awareness (SA) and cognitive workload (Kaber, Perry, Segall, McClernon, & Prinzel, 2006; Parasuraman, Cosenzo, & De Visser, 2009), yet other studies measure machine outputs, such as error rates and scores (Glas, Kanda, Ishiguro, & Hagita, 2012; Prinzel, Freeman, Scerbo, Mikulka, & Pope, 2000).

Feeling pressured to get the evaluations done quickly, Tim decides to stick with just measuring error rates by recording the number of mistakes made while retrieving an object. After Tim's assessment, AAEC deploys the robot, and astronauts begin training with it for future missions. However, it immediately becomes clear that the astronauts have difficulty interacting with and using the robot. Astronauts have so much trouble teleoperating the robot that robot collisions with obstacles are common, often resulting in necessary repairs. AAEC eventually ends the project altogether, and Tim's first major project as an AAEC aerospace engineer is considered a complete failure.

This example may seem dramatic, but it is not an uncommon occurrence. Frequently, machines in all domains, not just in the spaceflight domain, are carefully designed according to established requirements but are not properly evaluated for performance and safety because of a variety of misconceptions, including the one mentioned in this scenario—if the designer can use it, anyone should be able to use it. Often these assumptions lead to inadequate evaluation of the machines, which in turn results in machines that cannot be operated safely or at peak efficiency by the intended operators. Oversight of adequate testing for something as simple as an on-off switch for a vehicle could cause death, as evidenced in several modern cases where keyless cars have been left on by mistake due to lack of indication that they were on. In many cases, this mistake has resulted in carbon monoxide poisoning (McCoppin & Berger, 2015). As convenient as these vehicles are, they are quiet and typically come with no indicators of being left on when occupants exit the vehicle—this danger has been clearly overlooked in the design of such vehicles.

Research from several domains, including aerospace, has highlighted challenges in implementing human–machine systems—especially as machines have become increasingly automated—because these partnerships can elicit risky conditions leading to negative consequences (e.g., Endsley & Kaber, 1999; Lee & See, 2004; Parasuraman & Manzey, 2010; Parasuraman, Sheridan, & Wickens, 2000). For example, human–machine interaction shortcomings were described as contributing to the collision of *Progress 234* with the Mir space station (Billman, Feary, & Rochlis-Zumbado, 2011; Ellis, 2000). Therefore, it is important for system designers, operators, and key personnel to assess the safety and performance of these machines to optimize their functioning in the work environment. This concern includes the need to establish criteria for valid and reliable measurement of safety and performance in human–machine systems. However, in addition to directly examining safety and performance in human–machine systems, we must also consider and assess the factors *influencing* safety and performance. The first step to measurement is determining *what* needs to be measured, which is precisely

what we aim to show here. We will do so by highlighting categories of factors that should be considered for assessment in human–machine systems as well as indicating relationships that may exist between these categories.

So, what needs to be measured? Although it is easy to say "SA affects performance" or "workload affects safety," it is often much more difficult to see the larger picture: that these variables and many others converge to affect both safety and performance. Thus, in order to identify relevant factors influencing safety and performance that should be measured, we have created a guiding framework that considers how various human, machine, and contextual inputs, as well as human states and processes, intricately relate to safety and performance. The framework developed here highlights key areas of consideration for assessment in human–machine systems. Although other frameworks for the measurement of human–machine systems exist (Olsen & Goodrich, 2003; Pina, Cummings, Crandall, & Della Penna, 2008; Steinfeld et al., 2006), several limitations should be noted. For example, although some frameworks provide valuable information about the measurement of several human–machine system constructs (Steinfeld et al., 2006), they do not take into account the existence of precursors to the interaction itself. For example, without taking into account biases that humans bring to the interaction, engineers may create systems that humans do not want or understand how to use.

In other frameworks that do take into account the existence of precursors (Pina et al., 2008), the complexity of relationships between variables being measured is not adequately considered. Without considering how various inputs converge in human–machine interaction, one may misunderstand the full process by which a safety breakdown may occur and thus may miss an indicator of failure. The proposed framework is unique in that we consider not only necessary precursors but also the complexity of their interaction in affecting safety and performance. The framework allows researchers and practitioners to examine human–machine systems with a level of granularity that makes it easier to identify indicators of system failure while also increasing the likelihood of success of machine design for future human–machine systems.
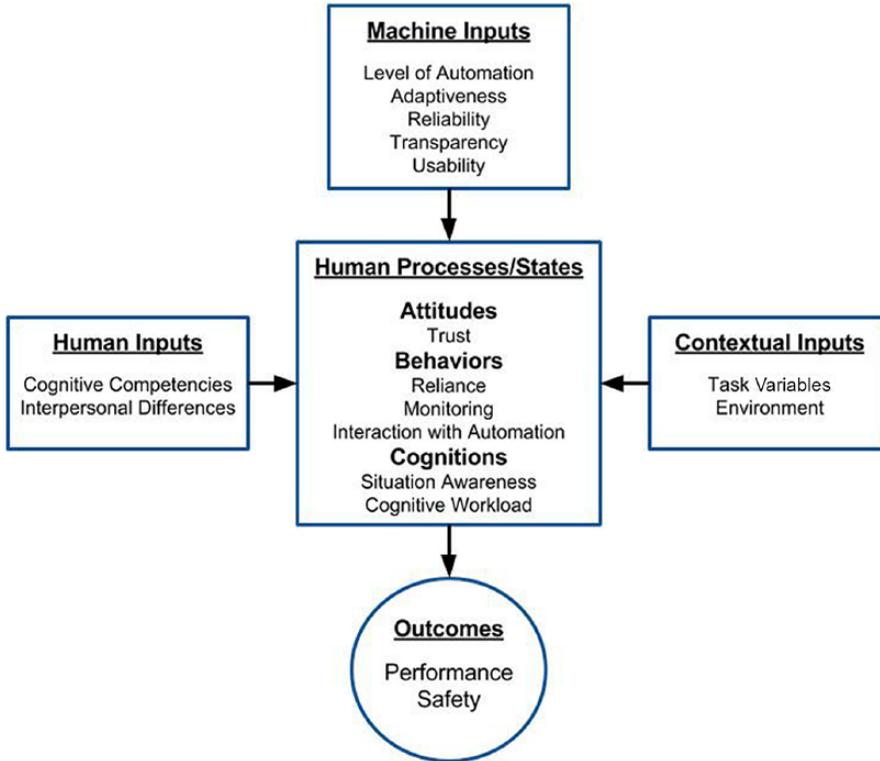
Figure 1. Framework of factors affecting safety and performance in human–machine systems.

We aim to add to the existing body of knowledge by considering measurable factors that empirical and theoretical research suggests may affect safety and performance in human–machine systems. As such, we have attempted to build our framework to encompass the full state of the science of human–machine systems. The state of the science, as it currently stands, includes a large focus on automation and autonomy. As such, many of the variables and relationships identified in this framework are most readily applied to systems that include automated and autonomous machines. Next, we explore the proposed framework in detail by first discussing its underpinnings followed by a description of each of the components of the framework. In order to maintain necessary brevity and focus on the outcomes of interest, we will go into detail only about precursors that the literature finds to be most widely associated with performance and safety outcomes. Finally,

we will discuss practical implications and limitations of the framework.

FRAMEWORK OVERVIEW

Figure 1 presents a framework explicating facets of modern human–machine systems that should be considered for measurement—particularly in relation to the interaction of a single human and one or multiple machines in such systems. This framework identifies and presents several constructs as they relate to two key outcomes of interest: safety and performance. We identified and used these constructs as part of an in-depth qualitative review (Oglesby, Iwig, Stowers, Sonesh, Leyva, & Salas, 2017). Only peer-reviewed, empirical articles were included in the analysis. We categorized each article by which variables were examined experimentally, giving a snapshot of the state of science on variables of consideration in human–machine systems. The results

subsequently affirmed the variables chosen in our final framework.

As can be seen in the figure, inputs to human–machine systems have been divided into three categories: human, machine, and contextual. Human processes and states link these inputs to the key outcomes and are divided into attitudinal, behavioral, and cognitive variables. For the purpose of this framework, which is meant to inform each near-term interaction between single operators and machines, safety and performance outcomes are reached through a combination of these aforementioned inputs, processes, and states. This framework is not meant to convey all of the ways in which safety and performance may be affected by specific variables, nor is it meant to convey every single variable that can affect these outcomes. Instead, it serves as a high-level snapshot of the state of the science, which typically examines human–machine interaction as being somewhat represented by input–process–outcome patterns.

As we describe the key components of the human–machine system framework throughout this article, we will be detailing the framework backward, starting with outcomes and working toward inputs. By doing so, we can keep the end goals of human–machine systems in mind while exploring various antecedents to such intended outcomes. Specifically, we first explore the outcomes that effective human–machine systems must achieve (i.e., safety and performance). Second, we describe human attitudinal, behavioral, and cognitive variables that serve to facilitate the achievement of these outcomes. Last, we specify three broad classes of inputs (i.e., human, machine, and contextual) that serve to directly affect human states and processes. We include brief discussions on the importance of key facets of human–machine interaction in obtaining optimal outcomes. Although this framework, for the sake of simplicity, necessarily focuses on the most salient and well-supported variables that have emerged in the literature, we do our best to call attention to additional variables of interest that deserve more research. As such, gaps in this framework should be considered as areas in the state of the science that can be improved upon or examined further.

OUTCOMES

As previously discussed, the framework presented in this paper explores two primary outcomes that are important to consider in human–machine systems: safety and performance. In the following paragraphs, we define these variables and discuss their position as the anchor of our framework. In-depth exploration of these variables can aid in the selection of appropriate metrics for human–machine systems, especially when one considers their position in relation to the inputs, states, and processes that ultimately precede them in human–machine interaction. For example, an in-depth understanding of safety requirements and risks can help one decide if safety should be measured directly (e.g., by counting collisions and near misses) or indirectly (e.g., by measuring identified precursors that may be linked to safety breaches).

Safety

It is critical to understand how one can prevent accidents, errors, and injuries in human–machine systems. To do so, one must first operationalize safety in the context of human–machine systems. Safety can be operationalized as the lack of catastrophic accidents or other negative incidents (Mustafiz, Sun, Kienzle, & Vangheluwe, 2008), with the level of safety decreasing as the accident rate increases. We further characterize safety as performance (e.g., successful completion of tasks) and efficiency (e.g., timeliness) in safety-critical environments. That is, if optimizing safety is a goal of a system, then performance relating to that goal may be considered a measurement of safety itself. For example, performance in air traffic control involves, in part, the need to maintain appropriate spacing between aircraft guided to land. This performance goal is also safety critical, making it a characteristic of safety in this context.

Safety is paramount in exploration of isolated, confined, and extreme environments; in fact, it is non-negotiable. For example, although human space flight is known to be inherently risky, without ensuring that astronauts and the vehicles and systems they use are as safe as reasonably possible, safe exploration is not possible. As such, the

National Aeronautics and Space Administration (NASA) institutes extremely rigorous requirements for rating spacecraft standards of safety (NASA, 2016). However, despite the best and brightest minds coming together to proactively identify and prevent every possible failure, hazard, or risk, there have still been tragic accidents (e.g., *Apollo 1* in 1967, space shuttle *Challenger* in 1986, space shuttle *Columbia* in 2003).

So what else can be done to ensure safety within human spaceflight? Human factors psychology has long been invested in studying how to maximize machine safety (Billings, 1997; Hoc & Amalberti, 2007; Morel, Amalberti, & Chauvin, 2008; D. Woods, 2010; D. Woods, Johannesen, Cook, & Sarter, 1994). Often the answer is not only to maximize the automated machine but to look to other elements as part of a broader sociotechnical system. Relevant elements may include poor machine design, lack of training, and misuse of the machines by the user. Any number of these elements may combine and lead to disastrous outcomes, illustrating that safety is often not in the control of just one element of a human–machine system (Reason, 1990). It has thus been acknowledged that the organizations, subteams, designers, and operators meant to advance, improve, and regulate system safety often contribute to system breakdowns and potentially deadly accidents (Dekker, 2004). Theoretical and applied research on these topics has led to initiatives such as resilience engineering, or the design of systems to perform in a resilient manner that improves safety outcomes (Hollnagel, Woods, & Leveson, 2007).

One approach to improving safety lies in the identification of hazards and risks that may have been ignored due to being perceived as insignificant. NASA uses this approach in its accident precursor analysis to identify and analyze errors in the system before they become accidents (NASA, 2011). This robust analysis offers a starting point from which to consider other measures of safety. In addition to identifying errors in this manner, it is also prudent to take into account variables stemming from the human, machine, and environment that may predict or affect safety. Many of these variables may influence safety or aid in the avoidance of a safety breach. With the present framework we consider multiple inputs, processes, and states that may contribute to safety in human–machine systems.

Performance

We operationalize performance in this framework as the successful completion of goals and subgoals in human–machine systems. A conspicuous goal in human–machine systems is to achieve a level of performance beyond what a human can achieve alone. Today's goal of placing humans on Mars by the 2030s (Daines, 2015) is achievable only through such systems. Only through the combination of highly complex automated machines and human inputs of engineers, astronauts, and human operators can this lofty goal be achieved. However, this goal, most likely the most challenging mission undertaken by man, can be achieved only by ensuring that each subgoal and task that human–machine systems are meant to achieve are in fact successful. A high level of performance for many activities and tasks that serve as stepping stones to achieving the larger goal of reaching Mars, or any human–machine system goal for that matter, is critical.

Performance can also be more specifically explored in terms of efficiency, or the successful completion of a task with minimal time and effort spent by the human–machine system. By this definition, a successfully performed task is efficient only if completed in a timely manner and within desired limits of effort. Efficiency can be assessed by evaluating the time and effort spent by the human or the machine apart from the overall system (Pina et al., 2008).

It is important to recognize that our general definition of performance is contingent upon the goals themselves as well as inputs, processes, and human states that may affect performance concerning those goals. Furthermore, several factors may influence the efficiency of human–machine systems. It is thus important to consider the variables that interact to affect performance. This framework aims to show variables that affect performance both directly and indirectly by discussing three classes of human states and processes as well as three classes of inputs that influence performance.

HUMAN PROCESSES AND STATES

Human error is not random. It results from basic human mental abilities and physical skills combined with the features of the tools being used, the tasks assigned, and the operating environment. (Leveson, 2011, p. 273)

In our human–machine system framework, *human states* refers to cognitive, motivational, and affective components (Marks, Mathieu, & Zaccaro, 2001). *Human processes*, on the other hand, is defined as actions that convert inputs to outcomes through cognitive or behavioral activities (Marks et al., 2001). Human processes and states not only influence the outcomes of safety and performance but also can be affected by other preexisting factors. Thus, measurement of these variables is highly informative and necessary to achieve successful human–machine interaction.

On the basis of information in the human–machine interaction literature, we break down human processes and states into attitudes, behaviors, and cognitions, where attitudes represent what the user feels, behaviors represent what the user does, and cognitions represent how the user thinks. This particular breakdown of variables is based on two assumptions regarding how humans interact with their surroundings. First, individuals' *behaviors* are based on preexisting *attitudes* (Ajzen & Fishbein, 1980; Householder & Greene, 2003). Second, human *cognition* is separate from, and additionally influences, the ability to have intelligent *behavior* (Aizawa, 2015; Rupert, 2013). These three considerations are often ignored by system designers. However, these three attributes are great contributors to the likelihood that the human–machine system will succeed or fail. As such, it is essential for attitudes, behaviors, and cognitions to be continuously monitored and quantified, as they provide great insight into how to improve or design more usable, safe, and efficient systems.

As discussed next, a primary attitude that has been found to affect human–machine systems is human trust in automation. Salient behaviors include reliance, monitoring, and interaction with automation. Cognitive variables include constructs such as SA and cognitive workload.

All of these have been identified as antecedents of human–machine performance and safety. Assessment of all of these factors can provide a proactive way to anticipate safety and performance breaches before they happen. Thus, a detailed exploration of these factors offers a critical first step in understanding their role in human–machine systems as well as the necessity of their measurement.

Attitudes

In this framework we explore trust as the focal attitude prevalent in human–machine systems. Traditionally, trust has been explored at the intersection of complex human interactions (Mayer, Davis, & Schoorman, 1995). Trust also occurs at the intersection of human, machine, and contextual inputs and involves the belief of helpfulness of both the machine and other team members in the human–machine system. According to this definition, the operator experiences trust in the machine when he or she believes it will help him or her achieve a task goal. Trust in human–machine systems can also be conceptualized as the belief that agents, either automated or human, will help the human in uncertain situations (Lee & See, 2004). Trust has been identified as a key contributor to acceptance in many machines and the conviction with which human operators will make decisions based on information presented by the machine (Hancock et al., 2011). This construct becomes increasingly important as new machines are introduced. For example, the success of systems involving automated cars, which are becoming increasingly popular today (Meister, 2015), will rely heavily on operator trust. On the other hand, it is also important to consider that societal adaptation to the use of machines, including through training and even simple exposure to machines, can create learning effects that may improve human understanding and use of machines, subsequently influencing trust.

Trust is typically influenced by the reliability (discussed later; de Visser & Parasuraman, 2011), validity (de Vries, Midden, & Bouwhuis, 2003), and timeliness (Abe & Richardson, 2005) of the machine. Individual differences, such as age and culture, also play a role in trust (Ho, Wheatley, & Scialfa, 2005; Rau, Li, & Li, 2009).

When the user trusts the machine, the user may expect higher machine reliability and fewer errors (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003). It is therefore important for these expectations to be upheld in order to maintain user trust. It is also important to ensure that trust is not so high as to induce user overreliance and complacency (Lee & See, 2004; discussed in the next section).

To this end, it is important to consider the calibration of trust in human–machine systems so that safety and performance are optimized. The goal of trust calibration is for users of machines to learn an appropriate level of trust based on the performance of the automated machines with which they are interacting (McBride & Morgan, 2010). Appropriately calibrated trust can lead to appropriate reliance on machines (discussed next), thereby enhancing safety and performance or, at the very least, preventing disasters (Gao & Lee, 2006).

Behaviors

Reliance, monitoring, and interaction with automation are closely related behaviors. They also build off trust, which is expected if one considers that people's attitudes often direct their behavior. Each of these behaviors is an important variable affecting safety in human–machine systems, as they can significantly affect an operator's ability to appropriately react in safety-critical environments. We begin with a discussion of reliance and characterize monitoring and interaction with automation as they relate to reliance.

Reliance can be defined as the level of dependence on automation (Wood, 2004). Whereas trust is considered to be an attitude toward an automated machine, reliance is a behavior (Lee & See, 2004). Although trust often influences reliance (Merritt, Huber, LaChapell-Unnerstall, & Lee, 2014), these two variables are not mutually inclusive. For instance, a person can trust that a machine will complete a task but not rely on it to do so. In contrast, a human may not trust the machine but feel forced to rely on it due to other circumstances, such as high workload.

Too much reliance can lead humans to not use automated machines as intended—this effect

has also been characterized as "improper use," which may come in the form of automation disuse, misuse, or abuse (see Dzindolet, Pierce, Beck, & Dawe, 1999; Parasuraman & Riley, 1997). Overreliance can also lead to complacency, which is defined as the wrongful assumption that the machine is functioning correctly resulting in a lack of vigilance (Billings et al., 1976, as cited in Parasuraman & Manzey, 2010). Complacency results in lack of monitoring or paying attention to the automated machine (Sheridan & Parasuraman, 2005) and, subsequently, loss of overall SA (Endsley, 1996). Decrements in monitoring and SA elevate the risk that operators will fail to detect and manage machine failures that arise in a timely manner, thus increasing the potential for negative safety outcomes (Bahner, Hüper, & Manzey, 2008). Underreliance on automated machines can have an equally negative effect. For example, frequent false alarms in various machines can cause users to ignore critical indicators, causing accidents to occur (Parasuraman & Riley, 1997).

Similar to trust, a lot of thought has gone into the question of how much reliance is enough, and how much is too much, for optimizing safety and performance in human–machine systems. Due to safety concerns with the extremes of reliance, it is recommended that reliance on automated machines be kept at a moderate level, thereby allowing users to trust the machines at an appropriate level for detecting errors while enhancing performance, consistent with the trust calibration research presented earlier. The importance of maintaining appropriate levels of these factors underlines the necessity of their measurement, making them key points of consideration in the overall assessment of human–machine systems.

Cognitive Variables

Many cognitive variables have emerged as being important for human–machine system and should be considered integral to the study of human–machine systems. However, it can often be unclear where these highly complex factors exist on the continuum of physical, affective, and cognitive states. Furthermore, their relevance is often only appreciated in prolonged

interaction with machines or intense environments. Other factors such as SA and cognitive workload have been directly operationalized as cognitive variables that arise in a high variety of contexts and have thus emerged as clear cognitive indicators of performance and safety across many different types of human–machine systems. We cover these factors as primary variables in our framework, as there is an overabundance of research supporting the need for their assessment (Kaber et al., 2006; Parasuraman et al., 2009; Sauer, Nickel, & Wastell, 2013). Yet, we still urge researchers to consider other relevant factors that may emerge in specific situations but that are not part of the scope of this framework.

SA is defined as the perception of one's surroundings, the comprehension of their meaning, and ability to predict their future states (Endsley, 1988). SA is thus characterized as existing in three levels: perception, comprehension, and projection (i.e., projection of future states based on current knowledge; Endsley, 2000). SA is an important mediator of performance for both individuals and teams (Endsley, 2000; Salas, Prince, Baker, & Shrestha, 1995) and plays a pivotal role in understanding performance outcomes in human–machine system contexts. According to Endsley (1996), losses in SA result in human out-of-the-loop performance. In other words, degraded performance is linked to the operator lacking control, appropriate skill, or full awareness of the machine's automation (Endsley, 1995). To illustrate, Endsley and Kiris (1995) reported that participants in manual conditions developed a more complete understanding of the system state than those in fully automated conditions and in turn exhibited higher performance.

Similar studies have echoed the positive relation between SA and performance in human–machine system contexts. Endsley and Kaber (1999), for example, demonstrated higher human–machine system performance when a human is introduced to automation at the implementation of a process, thereby increasing SA and keeping the human in the loop. They also suggested that reduced opportunities to control the machine would result in inefficient performance and reduced success with failure recovery (Kaber & Endsley, 1997). Such reductions in performance due to lack of control or interaction

with the machine, essentially known as skill decay (Arthur & Bennett, 1998), can negatively affect safety, especially when faced with time-sensitive, safety-critical tasks. These findings are especially relevant for human–machine system safety and performance, as automated machines require that operators ascribe meaning to them to detect and predict potential problems.

Cognitive workload is also an important predictor of human–machine system safety and performance (Langan-Fox, Canty, & Sankey, 2009). Conceptualized as the relationship between resources demanded by a task and resources available to the operator to complete the task (Parasuraman, Sheridan & Wickens, 2008), cognitive workload is supposed to be reduced by automated machines. However, Miller and Parasuraman (2007) noted that both low and high levels of automation can place workload burdens on the operator. For instance, machines may require high involvement from the operator, and the operator may react by not using the machine at all or engaging in adaptive strategies, such as attending to another task. In fact, task adaptive strategies during periods of high cognitive workload may be an important indicator of cognitive overload in an operator (Kirlik, 1993; Parasuraman & Hancock, 2008), thus highlighting the important connection between cognitive workload and behavioral interaction with machines.

The amount of cognitive workload experienced by an operator is affected by many things, including the level of automation (LOA) of the machine (see Machine Inputs section) and task(s) they are required to execute (see Contextual Inputs section). Additionally—in line with cognitive workload theory—environmental context, including task operations and individual differences, act as primary sources of stress (Conway, Szalma, & Hancock, 2007; Hancock, Ross, & Szalma, 2007) influencing workload (Hancock & Warm, 1989). Operators who experience sustained periods of high cognitive workload may be less effective at responding effectively to future adverse events (Parasuraman et al., 2008), thereby threatening safety. Likewise, operators who experience sustained periods of exceptionally low workload, or underload, may also become so bored or distracted that they fail

to respond effectively to changing events, a concern of particular interest to long-term missions, particularly, NASA's planned trip to Mars (Oglesby & Salas, 2012). By using sound measures, practitioners can determine what levels of workload are acceptable for a task without compromising the safety or performance of the overall system.

INPUTS

According to the proposed framework, three categories of inputs affect human–machine systems: (1) human inputs, or factors initiated and changed by the human operator(s) in the system; (2) machine inputs, or factors initiated and changed by the machine itself; and (3) contextual inputs, or factors external to human and machine control. These three categories are adapted from Hancock and colleagues' (2011) categories affecting trust development in human–robot interaction. We argue that these categories affect more than trust development; they also interact to affect all human processes and states that lead to safety and performance. Their measurement is critical to the design of human–machine systems, including team member selection and machine interface design. Understanding how these factors influence human–machine systems offers an important first step in the selection of metrics and guidelines, which will guide the design of human–machine systems.

Human Inputs

> Virtually all systems contain humans, but engineers are often not taught much about human factors and draw convenient boundaries around the technical components, focusing their attention inside these artificial boundaries. (Leveson, 2011, p. 175)

Many human characteristics affect overall human–machine systems. Two variables that are not included in our framework, but that are important to note, include sex and age. Age, in particular, is a necessary consideration in the examination of any technologically based system. We do not include these variables here, as the impact of age and sex on human–machine systems is still underresearched and evolving, making it difficult to give specific recommendations for the optimization of safety and performance. However, we encourage researchers to consider these variables as needed in specific situations that warrant their exploration.

We have combined and organized the most thoroughly researched and impactful human characteristics into the following variables within our framework: cognitive competencies and interpersonal traits. We discuss cognitive competencies and interpersonal traits in detail next, with a focus on propensity to trust and personality as key traits of consideration.

Cognitive competencies. Cognitive competencies encompass such characteristics as prior experience, expertise, skill, spatial ability, and working memory. We conceptualize prior experience in a broad way, including prior interactions with other machines, which may affect interpersonal traits (discussed later). On the other hand, expertise represents mastery in a domain, whereas skills can be acquired and improved. Skills may be defined as levels that can fluctuate, whereas expertise is considered an achieved state with more stable skills (Bril, Rein, Nonaka, Wenban-Smith, & Dietrich, 2010). Skill and expertise can interact with other inputs, thus leading to various processes and outcomes. For example, skill decay (loss in skill for manual control) can occur if the user is highly reliant on an automated machine (Parasuraman et al., 2000). Additionally, expertise and skill associated with interacting with machines have been found to lead to a higher detection of automation failures (Parasuraman & Manzey, 2010), which is important for the safety and performance of human–machine systems. As such, it is important to quantify these skills, levels of expertise, and experience to ensure that the operator is equipped and "ready" to interact with the machine.

Spatial ability is a facet of intelligence composed of multiple abilities, such as recognition and manipulation of objects in multiple dimensions (Lathan & Tracey, 2002). Spatial ability has been correlated with greater effectiveness and accuracy in scanning for, detecting, and targeting objects in a simulated robot operation task (Chen & Barnes,

2012; Chen, Durlach, Sloan, & Bowens, 2008) and may also predict speed at completing an assigned robotic route (Chen et al., 2008; Lathan & Tracey, 2002). Spatial ability is an important skill for robotic teleoperation by astronauts and can be improved by NASA robotics training (Liu, Oman, Galvan, & Natapoff, 2013). Likewise, working memory is important to human–machine systems due to its relationship to workload (Steinfeld et al., 2006). All of this evidence indicates that spatial ability and working memory are important facets of an operator's cognitive competencies that influence processes and outcomes in human–machine systems.

Interpersonal traits. Interpersonal traits are individual differences in human tendencies that guide behavior (Wiggins, 1979). They include traits such as personality and propensity to trust. The American Psychological Association (2016) defines personality as individual differences in thoughts, behaviors, and feelings. A user's personality characteristics can affect the human–machine system in a number of ways. Preexisting personality characteristics of an operator may affect a user's performance when interacting with machines, for example, by affecting how the operator copes with workload demands (Szalma & Taylor, 2011).

Personality may also affect processes involved in human–machine interaction. For example, in a detailed study examining individual differences in interaction with automated machines, Szalma and Taylor (2011) found that neuroticism impairs processes such as working memory and sustained attention. They also found that someone high in extraversion is typically able to perform a task that has low reliability or LOA; yet this individual's performance would suffer with a highly reliable system due to greater onset of complacency. Conscientious individuals are able to perform well in a highly reliable system (Szalma & Taylor, 2011), whereas agreeable individuals excel at determining the appropriate level of trust (Lee & See, 2004; Szalma & Taylor, 2011). Finally, those high in openness to experience are more likely to check the automated machine for accuracy and less likely to blindly trust that it is correct (Szalma & Taylor, 2011). Given these findings, we can see that individual personality differences may predict interaction with machines and, by extension, safety and performance.

Propensity to trust, or the "general willingness to trust others" (Mayer et al., 1995, p. 715), is an important individual difference that relates to personality in that it does not change simply as a result of a set of interactions with an automated machine. For example, propensity to trust does not vary as a result of interacting with a trustworthy machine or person. It is considered a stable trait, whereas level of trust (as an attitude) itself can change depending on the characteristics of the trusted party. Indeed, some consider propensity to trust to be a personality trait, influenced by past experiences, that is subsequently used in identifying similar situations in which to apply or remove trust (Rotter, 1971). Researchers have found that individuals who have low propensity to trust characterize a computer character as less credible compared with those with higher propensity to trust (Cowell & Stanney, 2005). This perceived lack of credibility may affect the user's decision to rely on the machine, thus influencing the user's interaction with it and, ultimately, the overall safety and performance of the human–machine system.

Machine Inputs

Identifying only operator error or sabotage as the root cause of the accident ignores most of the opportunities for the prevention of similar accidents in the future. (Leveson, 2011, p. 28)

Machine inputs are traits of the machine that can affect its interaction with humans. Such traits can also affect both user decisions and how the user interacts with the machine. The influence of machine inputs on human–machine interaction is an important factor to consider in machine design, making it a key tenet to our framework. Although many machine characteristics may be considered, including size, anthropometry, and relationship to the user, we have selected five characteristics that are most impactful and have the most theoretical and empirical support in the literature. These machine characteristics are LOA, adaptiveness, reliability, transparency, and usability. It is already widely known that LOA and reliability are important (see Parasuraman et al., 2000;

Sheridan & Parasuraman, 2005). For example, the LOA, or degree to which the machine can complete tasks without human input, affects how the human interacts with the machine (Parasuraman et al., 2000). Similarly, reliability, or the consistency with which a machine completes tasks, can lead to the success or failure of the machine and the overall system (Sheridan & Parasuraman, 2005). In the increasingly dynamic and complex environments where human–machine systems currently exist, adaptiveness and transparency may become still more important yet. As such, we focus on these two variables here.

Adaptiveness. Adaptive automation (AA) is the dynamic allocation of control to an operator in order to increase overall system performance (Kaber, Wright, Prinzel, & Clamann, 2005). AA enables the operator to take control of the machine whenever the need for operator input arises (such as in a safety crisis). This characteristic is different from adaptability in that adaptiveness is machine controlled, whereas adaptability is human controlled (Chou, Lai, Chao, Lan, & Chen, 2015). More adaptive task allocation can increase monitoring processes of machines, potentially combating the threat of overreliance and loss of SA in humans (Parasuraman, Mouloua, & Hilburn, 1999; Parasuraman, Mouloua, & Molloy, 1996). AA is also one means for confronting the challenges associated with balancing operator workload in human–machine systems. This concept adds flexibility to the automation, providing relief to the operator while keeping the operator in the loop.

Parasuraman and colleagues (1996) conducted a study with two different types of adaptive task allocation, a model-based framework for task allocation (LOA changed based on known models and did not vary by individual) and a performance-based trigger for task allocation (LOA changed based on score criteria). They found that the model-based technique was able to transfer control from the machine to the user and back to the machine to provide enhanced detection. On the other hand, some have advocated for considering the human state. For example, Kaber and colleagues (2005) suggested a need to focus on the operator and what stressors these adaptive automation pose in

order to offer adaptive automation in a consistent, reliable manner.

Transparency. Transparency is the ability of an interface to inform an operator of the intent, reasoning, and future plans of a machine (Chen et al., 2014) and can be broadly conceptualized as levels of information communicated to the user by the machine. This capability is important because it can improve operator SA, whereas lack of transparency can be detrimental to SA, safety, and overall performance (Sarter, 1995). For example, increased transparency can provide a means for the machine to let the operator know what it is doing and why—in order to prevent the operator from unnecessarily overriding actions out of uncertainty or distrust.

Communication from the machine can take the form of visual or auditory feedback to alert the user of changes or failures, configuration of displays to provide the user with current machine state information, or any other communication of information indicating the level of performance of the machine at any given time. Assessing how the machine provides information to the user is important, as factors such as modality and rate of communication to the user can influence safety. For instance, the presence of haptic feedback is particularly effective in helping avoid collisions in machine operation (Lee & Kim, 2008). At the same time, the effectiveness of such transparency in machines depends on its usability—that is, its learnability, efficiency, memorability, errors, and corresponding user satisfaction (Nielsen, 1994). For example, although researchers recommend that being transparent about uncertainty in the machine can be beneficial and necessary (Endsley, 2011), some have also found that it can introduce usability issues (Stowers et al., 2016), thereby negating its potential benefits.

Contextual Inputs

Some factors affecting human–machine systems may be external, or contextual, in nature. In this framework, we consider anything that is not within the immediate control of a human or machine a contextual factor. It is important to take contextual inputs into consideration when measuring human–machine systems because they dictate the type of task being performed as well as the constraints being imposed on

the task. Because many contextual inputs are specific to an environment or task and can vary significantly, we limit this section to factors that are important across several contexts and have the most theoretical and empirical support in the literature to influence humans' interaction with their surroundings.

Task variables. Multiple task variables may affect performance and safety in human–machine systems, including multitasking, task type, task load, and task complexity. These variables may affect human–machine systems by interacting with other inputs to the system. For example, in a multitasking environment, the user is required to switch between tasks, which can decrease performance (Cullen, Rogers, & Fisk, 2013). The LOA of the machine may interact with the multitasking environment and determine how effective the user is at multitasking. Similarly, multitasking in an environment where the automation is not fully reliable can increase the operator's workload (Cullen et al., 2013).

Task load, or the number of resources or demands an operator has responsibility for, has also been found to affect humans' use of machines through maladaptive workload and SA (Biros, Daly, & Gunsch, 2004; Skitka, Mosier, & Burdick, 1999). Having a secondary task can result in delayed response time when switching from acceleration to braking in a moving vehicle (Donmez, Boyle, & Lee, 2007). High levels of demands, such as having multiple robots, can lead to lapses in detecting situations that may require the operator's attention (Crandall, Cummings, Penna, & de Jong, 2011). On the other hand, too few demands can result in boredom and disengagement, which can also contribute to decrements in monitoring performance (D'Mello, Olney, Williams, & Hays, 2012). Both of these extremes can lead to safety and performance threats.

Task complexity, a well-researched but often unclear variable, has classically been investigated as (a) the number of elements included in the task, (b) the relationship between task elements, and (c) the evolution of this relationship over time (W. Woods, 1986). A more all-encompassing definition states that it is "the aggregation of any intrinsic task characteristic that influences the performance of a task" (Liu & Li, 2012, p. 559), with

task characteristics of interest including the clarity, quantity, and diversity involved in the task (for a comprehensive review, see Liu & Li, 2012). This definition holds fairly consistently with classic human–machine interaction literature, which typically represents task complexity as the mere number of displays being monitored (e.g., Molloy & Parasuraman, 1996).

Especially in spaceflight, where tasks require execution by individuals with high levels of expertise and experience, task complexity is highly relevant. Campbell (1988) suggested that complex tasks may lead to an increase in information overload and information diversity. Furthermore, having options can complicate a task when certain options lead to failure. Maynard and Hakel (1997) suggest that in addition to the actual complexity of that task, the *perception* of its complexity affects performance, whereby the more complex a task seems, the worse the performance will be.

Task complexity can also influence safety, depending on the context of the situation and other variables involved. For example, some groups have found that more difficult and complex tasks result in more collisions (e.g., Boessenkool et al., 2013). On the other hand, another study involving power plant operators showed that interpersonal characteristics played a role in the safety performance of workers in high-task-complexity conditions (Zhang, Ding, Li, & Wu, 2013). Taken together, these considerations beg the need for accurate measurement of the complexity of the task so that performance and safety countermeasures can be built in to moderate- and high-complexity tasks.

Environment. Environmental factors are conceptualized as characteristics of the human–machine system setting that are outside of the task, meaning that they are not influenced by the task, nor do they influence the task directly. Environmental factors are important to consider because, although they do not directly affect the task, they may have an impact on the operators' ability to complete the task. Specifically, environmental stimuli may interact with human and machine behavior, highlighting a need to consider environment in human–machine systems. Such stimuli may include noise, temperature, and altitude and

can affect the interaction between humans and their tasks, making it difficult to maintain awareness (Sarter, Woods, & Billings, 1997). For example, an airplane cockpit may change modes when a preset altitude is reached, without requiring action by the user. Additionally, factors such as noise have been shown to affect the LOA selected (Sauer et al., 2013). Like noise that distracts the user from the current task, other variables in the environment can play a significant role in operator task performance.

DISCUSSION

The framework we have outlined here serves as a tool for informing critical factors to measure by illustrating connections between categories of antecedents and outcomes that affect safety and performance in human–machine systems. By considering these factors and their relationships, one can take a multifaceted approach to measuring and monitoring safety and performance. Specifically, instead of taking the narrow approach of focusing only on the direct measurement of safety and performance, this framework can assist practitioners by encouraging a broader-spectrum approach to predict safety and performance by measuring the factors that precede their outcomes. For example, through the monitoring of factors discussed herein, practitioners may be able to detect degradations in performance. Where there are degradations in performance, a correction or calibration of one of the offending factors may lead to a correction in performance. Safety can be improved in a similar fashion.

Utilizing this framework when assessing human–machine systems will likewise enable designers to be more proactive in ensuring system success by using it as a predictive tactic for determining design requirements in human–machine systems. Specifically, designers may benefit most from considering how the machine inputs can be designed in such a way that they interact with human and contextual inputs to yield the best outcomes. For example, taking user population and interpersonal differences into account may help designers specify machine transparency for optimal safety or performance.

In the scenario presented in the opening of this paper, Tim was unsure how to evaluate the safety and performance of the robot and incompletely assessed these outcomes, which in turn led to the failure of the robot. Now let us imagine this scenario assuming that Tim has access to the framework presented here and uses it to guide his assessment of human–machine system safety and performance. In using this framework, Tim begins to understand the connections between human, machine, and contextual inputs as well as their impact on outcomes of interest. With this knowledge, he realizes that to assess safety and performance of his teleoperated robot, it is important that he manipulate environment and task variables during evaluation so they match the operational conditions as closely as possible. Additionally, he should assess the robot's adaptiveness and utility, the human's SA and cognitive workload when operating the robot, and outcome variables—such as operator satisfaction; errors, such as collisions or deviations from the preferred path; and task completion time. Now he has the information he needs to begin accurately assessing the safety and performance of the robot. The results of these assessments reveal specific areas that require improvement, which Tim addresses in the second prototype. In the end, Tim and his team are able to deliver a final prototype that the astronauts are able to safely and efficiently operate.

This example demonstrates the benefits of applying the theoretical framework presented in this paper toward developing guidelines and recommendations for measurement of human–machine systems. Clearly, human–machine system safety and performance are contingent on a variety of interrelated factors. We argue that it is important to consider each of these factors during assessment. Such consideration has many practical implications that can assist practitioners, designers, and machine users across many contexts. It is through consideration of these factors that the design and measurement of human–machine systems, as well as countermeasures to combat safety and performance decrements in such systems, can be improved.

Limitations and Next Steps

To keep this framework manageable and verifiable, we have not included all possible factors that may affect human–machine systems or

all possible relationships that may exist between factors. However, we included the factors that apply across most contexts in human–machine systems and have the highest known influence on human–machine systems. Additionally, although this framework guides the user on what should be measured when assessing safety and performance in human–machine systems, it does not offer advice specific to the type of method of measurement for each factor presented. For example, this framework does not show when factors should be measured or how they should be measured. It merely shows what should be measured, why such factors are important, and how they relate to each other.

Next steps should include not only the quantification of these factors through measurement but an overall verification of the framework itself. Specifically, the quantification of this framework via meta-analysis and other empirical methods will fine-tune the framework still more while also allowing researchers to more accurately utilize it as a guide. Additional work in this area includes the creation of a tool kit that compiles all of the information considered here such that designers and programmers can easily access metrics and measurement techniques that are most appropriate for assessing the safety and performance of their specific systems. Our literature analysis that informed the development of this framework creates a first step in doing this by compiling the relevant literature (Oglesby et al., 2017), but more work is currently in progress to create such a tool kit. Finally, the application of this framework to specific contexts should be explored. Although we have designed the framework to be broad and far-reaching, determining how it can be best utilized and followed in certain contexts (e.g., spaceflight) and tasks (e.g., robotic operation vs. monitoring tasks) will make it still more usable to specific situations.

With these next steps in mind, we expect that this framework can be very useful to the design and assessment of human–machine systems. By considering the multitude and complexity of variables informing safety and performance outcomes, it becomes possible not only to assess outcomes but to prevent failures through the use of countermeasures in the design and implementation of machines.

ACKNOWLEDGMENTS

KEY POINTS

- The first step to accurate measurement of human–machine systems is deciding what to measure and how such factors relate to each other.
- In order to accurately measure and predict performance and safety in human–machine systems, factors that affect these two outcomes must be considered.
- Inputs from the human, machine, and environment interact to result in several attitudes, behaviors, and cognitive variables, which ultimately shape performance in human–machine systems.

REFERENCES

Abe, G., & Richardson, J. (2005). The influence of alarm timing on braking response and driver trust in low speed driving. *Safety Science*, *43*, 639–654.

Aizawa, K. (2015). Cognition and behavior. *Synthese*. Advance online publication. doi:10.1007/s11229-014-0645-5

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behaviour*. London, UK: Pearson.

American Psychological Association. (2016). *Personality*. Retrieved from http://www.apa.org/topics/personality/

Arthur, W., Jr., & Bennett, W., Jr. (1998). Factors that influence skill decay and retention. *Human Performance*, *11*, 57–101.

Bahner, J. E., Hüper, A. D., & Manzey, D. (2008). Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human–Computer Studies*, *66*, 688–699.

Billings, C. E. (1997). *Aviation automation: The search for a human-centered approach*. Mahwah, NJ: Lawrence Erlbaum.

Billman, D., Feary, M., & Rochlis-Zumbado, J. (2011). *Evidence report: Risk of inadequate design of human and automation/robotic integration*. Retrieved from https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20110014661.pdf

Biros, D. P., Daly, M., & Gunsch, G. (2004). The influence of task load and automation trust on deception detection. *Group Decision and Negotiation*, *13*, 173–189.

Boessenkool, H., Abbink, D. A., Heemskerk, C. J. M., Steinbuch, M., de Baar, M. R., Wildenbeest, J. G. W., Ronden, D., & Koning, J. F. (2013). Analysis of human-in-the-loop tele-operated maintenance inspection tasks using VR. *Fusion Engineering and Design*, *88*, 2164–2167.

Bril, B., Rein, R., Nonaka, T., Wenban-Smith, F., & Dietrich, G. (2010). The role of expertise in tool use: Skill differences in functional action adaptations to task constraints. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 825–839.

Campbell, D. J. (1988). Task complexity: A review and analysis. *Academy of Management Review*, *13*, 40–52.

Chen, J. Y., & Barnes, M. J. (2012). Supervisory control of multiple robots: Effects of imperfect automation and individual differences. *Human Factors*, *54*, 157–174.

Chen, J. Y., Durlach, P. J., Sloan, J. A., & Bowens, L. D. (2008). Human–robot interaction in the context of simulated route reconnaissance missions. *Military Psychology*, *20*, 135.

Chen, J. Y., Procci, K., Boyce, M., Wright, J., Garcia, A., & Barnes, M. (2014). *Situation awareness-based agent transparency* (No. ARL-TR-6905). Aberdeen Proving Ground, MD: Army Research Lab Human Research and Engineering Directorate.

Chou, C. Y., Lai, K. R., Chao, P. Y., Lan, C. H., & Chen, T. H. (2015). Negotiation based adaptive learning sequences: Combining adaptivity and adaptability. *Computers & Education*, *88*, 215–226.

Conway, G. E., Szalma, J. L., & Hancock, P. A. (2007). A quantitative meta-analytic examination of whole-body vibration effects on human performance, *Ergonomics*, *50*(2), 228-245

Cowell, A. J., & Stanney, K. M. (2005). Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility. *International Journal of Human–Computer Studies*, *62*, 281–306.

Crandall, J. W., Cummings, M. L., Penna, M. D., & de Jong, P. (2011). Computing the effects of operator attention allocation in human control of multiple robots. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, *41*, 385–397.

Cullen, R. H., Rogers, W. A., & Fisk, A. D. (2013). Human performance in a multiple-task environment: Effects of automation reliability on visual attention allocation. *Applied Ergonomics*, *44*, 962–968.

Daines, G. (2015). *NASA's journey to Mars*. Retrieved from http://www.nasa.gov/content/nasas-journey-to-mars

de Visser, E., & Parasuraman, R. (2011). Adaptive aiding of human–robot teaming: Effects of imperfect automation on performance, trust, and workload. *Journal of Cognitive Engineering and Decision Making*, *5*, 209–231.

de Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning. *International Journal of Human–Computer Studies*, *58*, 719–735.

Dekker, S. (2004). *Ten questions about human error: A new view of human factors and system safety*. Boca Raton, FL: CRC Press.

D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human–Computer Studies*, *70*, 377–398.

Donmez, B., Boyle, L. N., & Lee, J. D. (2007). Safety implications of providing real-time feedback to distracted drivers. *Accident Analysis & Prevention*, *39*, 581–590.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human–Computer Studies*, *58*, 697–718.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (1999). Misuse and disuse of automated aids. In *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (pp. 339–343). Santa Monica, CA: Human Factors and Ergonomics Society.

Ellis, S. R. (2000). Collision in space. *Ergonomics in Design*, *8*(1), 4–9.

Endsley, M. R. (1988, May). Situation awareness global assessment technique (SAGAT). In *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference 1988* (pp. 789–795). New York, NY: IEEE.

Endsley, M. R. (1995). A taxonomy of situation awareness errors. *Human Factors in Aviation Operations*, *3*, 287–292.

Endsley, M. R. (1996). Automation and situation awareness. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 163–181). Mahwah, NJ: Lawrence Erlbaum.

Endsley, M. R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 131–152). Mahwah, NJ: Lawrence Erlbaum.

Endsley, M. R. (2011). *Designing for situation awareness: An approach to user-centered design*. Boca Raton, FL: CRC Press.

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness, and workload in a dynamic environment. *Ergonomics*, *42*, 462–492.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, *37*, 381–394.

Gao, J., & Lee, J. D. (2006). Extending the decision field theory to model operators' reliance on automation in supervisory control situations. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, *36*, 943–959.

Glas, D. F., Kanda, T., Ishiguro, H., & Hagita, N. (2012). Temporal awareness in teleoperation of conversational robots. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, *42*, 905–919.

Hancock, P. A., Billings, D. R., Oleson, K. E., Chen, J. Y., De Visser, E., & Parasuraman, R. (2011). *A meta-analysis of factors influencing the development of human–robot trust* (No. ARL-TR-5857). Aberdeen Proving Ground, MD: Army Research Lab.

Hancock, P. A., Ross, J. M., & Szalma, J. L. (2007). A meta-analysis of performance response under thermal stressors. *Human Factors*, *49*, 851–877.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, *31*, 519–537.

Ho, G., Wheatley, D., & Scialfa, C. T. (2005). Age differences in trust and reliance of a medication management system. *Interacting With Computers*, *17*, 690–710.

Hoc, J. M., & Amalberti, R. (2007). Cognitive control dynamics for reaching a satisficing performance in complex dynamic situations. *Journal of Cognitive Engineering and Decision Making*, *1*, 22–55.

Hollnagel, E., Woods, D. D., & Leveson, N. (2007). *Resilience engineering: Concepts and precepts*. Aldershot, UK: Ashgate.

Householder, H. J., & Greene, K. L. (2003). The theory of reasoned action. In J. P. Dillard & M. Pfau (Eds.), *The persuasion handbook* (pp. 259–288). Thousand Oaks, CA: Sage.

Kaber, D. B., & Endsley, M. R. (1997). Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress*, *16*, 126–131.

Kaber, D. B., Perry, C. M., Segall, N., McClernon, C. K., & Prinzel, L. J., III. (2006). Situation awareness implications of adaptive automation for information processing in an air traffic control-related task. *International Journal of Industrial Ergonomics*, *36*, 447–462.

Kaber, D. B., Wright, M. C., Prinzel, L. J., & Clamann, M. P. (2005). Adaptive automation of human-machine system information-processing functions. *Human Factors*, *47*, 730–741.

Kirlik, A. (1993). *Acquisition and production of skilled behavior in dynamic decision-making tasks*. Retrieved from https://ntrs.nasa.gov/search.jsp?R=19920007914

Langan-Fox, J., Canty, J. M., & Sankey, M. J. (2009). Human–automation teams and adaptable control for future air traffic management. *International Journal of Industrial Ergonomics*, *39*, 894–903.

Lathan, C. E., & Tracey, M. (2002). The effects of operator spatial perception and sensory feedback on human–robot teleoperation performance. *Presence: Teleoperators and Virtual Environments*, *11*, 368–377.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*, 50–80.

Lee, S., & Kim, G. J. (2008). Effects of haptic feedback, stereoscopy, and image resolution on performance and presence in remote navigation. *International Journal of Human–Computer Studies*, *66*, 701–717.

Leveson, N. (2011). *Engineering a safer world: Systems thinking applied to safety*. Cambridge, MA: MIT Press.

Liu, A. M., Oman, C. M., Galvan, R., & Natapoff, A. (2013). Predicting space telerobotic operator training performance from human spatial ability assessment. *Acta Astronautica*, *92*, 38–47.

Liu, P., & Li, Z. (2012). Task complexity: A review and conceptualization framework. *International Journal of Industrial Ergonomics*, *42*, 553–568.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review*, *26*, 356–376.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*, 709–734.

Maynard, D. C., & Hakel, M. D. (1997). Effects of objective and subjective task complexity on performance. *Human Performance*, *10*, 303–330.

McBride, M., & Morgan, S. (2010). *Trust calibration for automated decision aids*. Retrieved from https://www.ihssnc.org/portals/0/Documents/VIMSDocuments/McBride_Research_Brief.pdf

McCoppin, R., & Berger, S. (2015, June 18). Poisoning death of Highland Park couple points to alarming trend. *Chicago Tribune*. Retrieved from http://www.chicagotribune.com/suburbs/lake-county-news-sun/news/ct-keyless-ignition-risk-met-20150618-story.html

Meister, J. (2015). *Survey shows amount of trust in autonomous vehicles*. Retrieved from http://www.pddnet.com/news/2015/10/survey-shows-amount-trust-autonomous-vehicles

Merritt, S. M., Huber, K., LaChapell-Unnerstall, J., & Lee, D. (2014). *Continuous calibration of trust in automated systems* (AFRL-RH-WP-TR-2014-0026). Wright-Patterson Air Force Base, OH: Air Force Research Laboratory.

Miller, C. A., & Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human Factors*, *49*, 57–75.

Molloy, R., & Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors*, *38*, 311–322.

Morel, G., Amalberti, R., & Chauvin, C. (2008). Articulating the differences between safety and resilience: The decision-making process of professional sea-fishing skippers. *Human Factors*, *50*, 1–16.

Mustafiz, S., Sun, X., Kienzle, J., & Vangheluwe, H. (2008). Model-driven assessment of system dependability. *Software & Systems Modeling*, *7*, 487–502.

National Aeronautics and Space Administration. (2011). *Accident precursor analysis handbook*. Retrieved from https://www.hq.nasa.gov/office/codeq/doctree/NASA_SP-2011-3423.pdf

National Aeronautics and Space Administration. (2016). *NASA Technical Standards System (NTSS)*. Retrieved from https://standards.nasa.gov/nasa-technical-standards

Nielsen, J. (1994). *Usability engineering*. Amsterdam, Netherlands: Elsevier.

Oglesby, J. M., Iwig, C., Stowers, K., Sonesh, S., Leyva, K., & Salas, E. (2017). *Human–automation systems safety and performance: A qualitative review of measures and metrics*. Manuscript submitted for publication.

Oglesby, J. M., & Salas, E. (2012). The issue of monotony and low workload in spaceflight: Considerations for the mission to Mars. In *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting* (pp. 1782–1786). Santa Monica, CA: Human Factors and Ergonomics Society.

Olsen, D. R., & Goodrich, M. A. (2003). Metrics for evaluating human–robot interactions. *All Faculty Publications*, *23*, Paper 1052.

Parasuraman, R., Cosenzo, K. A., & De Visser, E. (2009). Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology*, *21*, 270.

Parasuraman, R., & Hancock, P. A. (2008). Mitigating the adverse effects of workload, stress, and fatigue with adaptive automation. In J. L. Szalma & P. A. Hancock (Eds.), *Performance under stress* (pp. 45–58). Aldershot, UK: Ashgate.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, *52*, 381–410.

Parasuraman, R., Mouloua, M., & Hilburn, B. (1999). Adaptive aiding and adaptive task allocation enhance human-machine interaction. In M. W. Scerbo (Ed.), *Automation technology and human performance: Current research and trends* (pp. 119–123). Boca Raton, FL: CRC Press.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors*, *38*, 665–679.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*, 230–253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, *30*, 286–297.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, *2*, 140–160.

Pina, P. E., Cummings, M. L., Crandall, J. W., & Della Penna, M. (2008). Identifying generalizable metric classes to evaluate human-robot teams. In *Proceedings of the 3rd Annual Conference on Human–Robot Interaction* (pp. 13–20). New York, NY: ACM.

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2000). A closed-loop system for examining psychophysiological measures for adaptive task allocation. *International Journal of Aviation Psychology*, *10*, 393–410.

Rau, P. P., Li, Y., & Li, D. (2009). Effects of communication style and culture on ability to accept recommendations from robots. *Computers in Human Behavior*, *25*, 587–595.

Reason, J. (1990). *Human error*. Cambridge, UK: Cambridge University Press.

Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist*, *26*, 443.

Rupert, R. (2013). Memory, natural kinds, and cognitive extension; or, Martians don't remember, and cognitive science is not about cognition. *Review of Philosophy and Psychology*, *4*, 25–47.

Salas, E., Prince, C., Baker, D. P., & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. *Human Factors*, *37*, 123–136.

Sarter, N. B. (1995). Knowing when to look where: Attention allocation on advanced automated flight decks. In *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 239–242). Columbus: Ohio State University, Department of Aerospace Engineering, Applied Mechanics and Aviation.

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 1926–1943). New York, NY: Wiley.

Sauer, J., Nickel, P., & Wastell, D. (2013). Designing automation for complex work environments under different levels of stress. *Applied Ergonomics*, *44*, 119–127.

Sheridan, T. B., & Parasuraman, R. (2005). Human–automation interaction. *Reviews of Human Factors and Ergonomics*, *1*, 89–129.

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human–Computer Studies*, *51*, 991–1006.

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006). Common metrics for human–robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human–Robot Interaction* (pp. 33–40). New York, NY: ACM.

Stowers, K., Kasdaglis, N., Newton, O., Lakhmani, S., Wohleber, R., & Chen, J. (2016). Intelligent agent transparency: The design and evaluation of an interface to facilitate human and artificial agent collaboration. In *Proceedings of the Human Factors and Ergonomics 60th Annual Meeting* (pp. 1706–1710). Santa Monica, CA: Human Factors and Ergonomics Society.

Szalma, J. L., & Taylor, G. S. (2011). Individual differences in response to automation: The five factor model of personality. *Journal of Experimental Psychology: Applied*, *17*, 71.

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, *37*, 395.

Wood, S. (2004). *Flight crew reliance on automation* (CAA Paper 10). Retrieved from https://publicapps.caa.co.uk/docs/33/2004_10.PDF

Woods, D. D. (2010). *Behind human error*. Aldershot, UK: Ashgate.

Woods, D. D., Johannesen, L. J., Cook, R. I., & Sarter, N. B. (1994). *Behind human error: Cognitive systems, computers and hindsight* (No. CSERIAC-SOAR-94-01). Wright Patterson Air Force Base, OH: Crew Systems Ergonomics Information Analysis Center.

Woods, W. A. (1986). Important issues in knowledge representation. *Proceedings of the IEEE*, *74*, 1322–1334.

Zhang, J., Ding, W., Li, Y., & Wu, C. (2013). Task complexity matters: The influence of trait mindfulness on task and safety performance of nuclear power plant operators. *Personality and Individual Differences*, *55*, 433–439.

Kimberly Stowers is a doctoral candidate in applied experimental and human factors psychology at the University of Central Florida. She received an MS in modeling and simulation from the same university. Her research focuses include the study of attitudes, behaviors, and cognition in human–machine interaction.

James Oglesby is a doctoral candidate at the University of Central Florida, where he received a BS in psychology. He is a research assistant at the Institute for Simulation and Training in Orlando, Florida. His research interests include team performance, team cognition, simulation and games for learning, and performance and habitability in extreme environments.

Shirley Sonesh obtained her doctorate in organizational behavior at the A. B. Freeman School of Business at Tulane University in 2012. Currently, she is an adjunct professor at Tulane University and conducts research on the topics of expatriation, coaching, teamwork, training, and human automation systems.

Kevin Leyva graduated from the University of Central Florida with a master's degree in modeling and simulation. He received a BS in psychology from the same university. He has experience researching topics in human–machine teaming and usability. His research interests include automation, ecological research, and design.

Chelsea Iwig obtained her first master's degree from Embry-Riddle Aeronautical University in human factors and systems in 2014. A year later, in 2015, she received a second master's degree from the University of Central Florida in modeling and simulation. Currently, she is a PhD student and graduate research assistant at Rice University, where she conducts research on the topics of automation, team training, and performance measurement.

Eduardo Salas is a professor and Allyn R. and Gladys M. Cline Chair in Psychology at Rice University. He has coauthored over 450 journal article and book chapters and has coedited 27 books. His expertise includes assisting organizations to foster teamwork, design and implement team training strategies, facilitate training effectiveness, manage decision making under stress, develop performance measurement tools, and create a safety culture.