ASSESSING HUMAN-AUTOMATION SYSTEM SAFETY, EFFICIENCY, AND PERFORMANCE: DEVELOPING A METRICS FRAMEWORK

James M. Oglesby, Kimberly Stowers, Kevin Leyva, Aaron Dietz, Shirley Sonesh, Shawn Burke, & Eduardo Salas
Institute for Simulation and Training, Department of Psychology, University of Central Florida

Automation is an important and widely utilized component in work environments across many domains; it is useful for completing tasks too dangerous or cumbersome for personnel to complete by themselves. Despite its benefits, potential issues can occur that may impact safety and efficiency in the overall human-automation system. To realize the benefits of automation, designers must be able to measure and assess the levels of safety and efficiency. This paper will discuss a theoretical framework to guide the development and selection of metrics for assessing human-automation interaction.

INTRODUCTION

Although automation is useful in a vast range of domains to help personnel complete tasks, potential issues can arise that could significantly impact the safety and efficiency in work environments that use these systems (Parasuraman, Sheridan, & Wickens, 2000; Endsley & Kaber, 1999; Lee & See, 2004; Parasuraman & Manzey, 2010). To identify potential threats to automation safety and performance, it is necessary to evaluate the contributing factors to automation performance. While there are established guidelines for designing automation systems for a variety of contexts (e.g., Goodrich & Olsen, 2003), system developers need to have the proper resources to measure the safety and efficiency of system designs for specific contexts and purposes.

This paper introduces a theoretical framework to guide the development of metrics for assessing human-automation system (HAS) safety and efficiency. The proposed framework is unique in that it attempts to capture the characteristics of both the human operator and the automation agentas potential causes of performance. Additionally, our framework considers the measurement issues associated with the complexity of relationships between factors influencing HAS performance.

Overall, this framework conceptualizes factors associated with HAS performance into three different categories: (1) inputs, (2) processes, and (3) outcomes. The impact of these categories on HAS performance, and how they impact measurement selection, will now be explained in detail in the following sections. (Figure 1).

INPUTS

Inputs are preconditions of HAS performance that are not necessarily dependent on the interaction between the user, the automation, or the environment. Input factors are divided into 3 different categories: (1) user inputs, (2) automation inputs, and (3) contextual inputs. User inputs pertain to factors that are focused on the human operator of the automation system, and can account for predicting how they will interact with the system under varying conditions. Automation inputs are associated with characteristics of the automation agent itself, which can determine how it operates to complete a task and with the user. Contextual inputs refer to factors not pertaining to either the user or automation agent, but assocated with environmental factors and task conditions (e.g. time pressure). These factors can influence how the operator and automation system interact with each other while conducting tasks.

User Inputs

Our theoretical framework identifies two main factors that are associated with human-automation interaction: personality and expertise. Personality—defined as "individual differences in characteristic patterns of thinking, feeling, and behaving" (American Psychologial Association, 2014)— affects a user's performance with an automation system. Specifically, Szalma and Taylor (2011) found that certain personality characteristics affect the way a person responds to tasks and affects their levels of workload and stress. For example, neuroticism impacts a user's process in adapting to changing task conditions or situations where the automation is affected such as during malfunctions (Matthews, Deary, & Whiteman, 2003). Additionally, personality dimensions such as 'openness'—defined as being open to novel situations and environments—may have more adaptability to operate an automation system under variable levels of reliability as they may be less likely to trust the system and continuously check its functioning (Szalma & Taylor, 2011).

Personality is typically measured using self-report methods that structure personality profiles into frameworks, such as the International Personality Item Pool (Goldberg, 2001), and is considered one of the most direct ways to assess individual personalities. In order to optimize the fit between a user's personality and the automation system, it is important for designers to consider how interventions (e.g., training) can help or hinder an operator's interaction and performance with a specified automation system.

Expertise is considered to be the user's knowledge and skill with a particular system, task, or domain (Ericsson & Lehmann, 1996). Expertise and skill with automationcan lead to greater detection of automation failures (Parasuraman & Manzey, 2010), even when expert operators indicate they have the same level of complacency as novices that interacted with the system. Being an expert in something generally relates to the persons interest in the system, past performance with the system or environment, or simply continued exposure to the system. It can be assessed through records of work history,

Copyright 2014 Human Factors and Ergonomics Society. DOI 10.1177/1541931214581240

training, certifications, and self report assessments of how long they have worked with a given system or domain.

Users develop expert knowledge and skills, based on knowledge of the domain and individual differences. Additionally, biases may be established based on their perception of the automation agent that can affect how operators interact with the system. These develop as mediating factors that indicate how user inputs affect the emergent processes and states within the HAS. For example, with propensity to trust—the willingness of an individual to rely on a person or agent to perform an action (Mayer, Davis, & Schoorman, 1995)—can lead to a likelihood of an over- or under-reliance on the automation to complete mission goals (Lee and See, 2004). The willingness to rely on an automation system can also be affected by expertise or previous experience. For example, if the operator's previous experience with a system has been mostly fixing errors made by the automation system, they would be less likely to trust the system to complete its task in future cases. Although some systems may be similar in function, the level of trust can vary if one system is familiar and the other is not (Szalma & Taylor, 2011). Due to history based influences in how operators' interactions with automation systems may change, metric across time can help in gaining insight into how these systems are used and possible projection for how they are used in future cases.

Automation Inputs

The design of an automation system is a critical component in predicting and ensuring the safety and efficiency of the entire system. If, for example, the system is not designed to allow the operator to intervene in case of automation failure, it can lead to accidents in the workplace (Parasuraman & Riley, 1997). Additionally, characteristics of the design that do not accommodate for the limitations of the user can also hinder performance. Characteristics associated with automation inputs include reliability, adaptability, and level of automation. In turn, these affect how the user interacts with the system and what decisions that user will make.

Reliability of the automation is defined as the ability of the system to perform as needed and for the length expected (Mustafiz, Sun, Kienzle, & Vangheluwe, 2008; Geffroy & Motet, 2002; Sheridan & Parasuraman, 2005). Reliability is important because not only do reliable systems provide better performance, users are also more likely to trust the automation (Wiegmann, Rich, & Zhang, 2001). However, problems emerge in high reliability systems as well. Over time, the user may rely too heavily on the accuracy of the automation and fail to detect errors when they eventually occur (Parasuraman & Riley, 1997). To counter the issues with reliability of automation, adaptability allows the automation to be dynamic. Adaptability is the ability of the automation to change the "division of labor between humans and machines" (Sheridan & Parasuraman, 2005, p.109). More adaptive task allocation can increase monitoring of automated systems, potentially combating the threat of complacency and loss in situation awareness (Parasuraman, Mouloua, & Molloy, 1996; Parasuraman, Mouloua, & Hillburn, 1999).

The level of automation of a system can be defined as the amount of control that the automation system has over the overall task (relative to the operator), and multiple frameworks exist that conceptualize how varying levels allocate different tasks to the operator and system (e.g., Endsley & Kaber, 1999; Parasuraman, Sheridan & Wickens, 2000). These types of frameworks allow one to categorize various systems based on their taskwork abilities as well as transparency of automation processes to the operator interacting with the automation. Therefore, these framework can aid in determining the properties of the automation system, and determine the projected role of the operator in regular taskwork, intervening for unique cases, and the operator's knowledge of the automation system's functioning.

Contextual Inputs

Contextual inputs are factors that exist within the environment or task that can affect how the human or system performs. Under this category, three main factors are considered: (1) team variables, (2) task complexity, and (3) the task environment.

Many operators work in teams and interact with the same system to complete tasks (e.g., nuclear power plant control room). Characteristics of the team—including team size—can can affect how team members use systems collaboratively (Huang & Hwang, 2009). Additionally, team performance is also affected by teamwork behaviors, such as communication with fellow teammates and collaborative behaviors (Fiore, Rosen, Smith-Jentsch, & Salas, 2010). In the cases where multiple operators interact with automation systems, it is important that measurements of human-automation interaction take into account team level variables that can impact safety and performance. Possible measurement considerations include team situation awareness (Burke, Lum, Scielzo, Smith-Jentsch, & Salas, 2009) and potential conflict and trust problems between team members when working together (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003).

Task complexity and environment are other contextual inputs associated with human-automation interaction that should be measured to determine if the system is ideally designed to work under specific conditions. Measurement implications of human-automation interaction should take into account task and environmental factors that can impact the overall system's process. For example, the time pressure on a task can affect the amount of reliance an operator has in an automation system to make decisions, regardless of whether they trust the automation system or not (Rice & Keller, 2009).

PROCESSES AND STATES

Major contributions to automation failures can be attributed to inefficient interactions between the operator and the automation system.Our theoretical framework identifies three categories of processes and states: (1) attitudes, (2) behaviors, and (3) cognitions. The following will provide information regarding the various factors that compose each of these categories, their relation to human-automation performance, and methods for measuring these factors.

Attitudes

Trust and complacency are two focal attitudes that affect human-automation interaction. Trust is defined as the belief that an automation agent will help the operator in varying circumstances (Lee & See, 2004), and differs from propensity to trust in that it is an emergent state from when the operator interacts with the automation rather than a stable trait. Complacency within high reliability systems may lower error detection (Oakley, Mouloua, & Hancock, 2003). This overreliance on the automation may also lead to reduced practice, misuse of automation, and skill decay (Parasuraman, Sheridan & Wickens, 2000).

Behaviors

Behaviors involved in HAS include monitoring and automation use. There are some cases where automation will fail to complete the tasks allocated to them effectively, requiring the operator to identify these failures and intervene to maintain performance. Monitoring involves tasks not necessarily relying on inputs of the operator, but on the operator ensuring the maintenance of an automation system during the task. Poor monitoring of automation and failure detection can be observed on both extremes of the trust spectrum. First, placing too much trust in automation results in complacency. This may lead to the operator decreasing attention to potential failures that may occur (Metzger & Parasuraman, 2001; Parasuraman, 2000). On the other hand, a lack of trust in the automation can also impact failure detection, as operators may tend to ignore alerts elicited by the automation when an incident occurs (Lee & See, 2004). There is evidence that additional factors of system design, operator characteristics, and the task itself can play a role in monitoring (Parasuraman & Manzey, 2010).

Automation use is defined as the voluntary actions conducted by the operator with the automation system (Parasuraman & Riley, 1997). This can encompass all of the interactions the operator conducts, ranging from proper actions and monitoring of the automation, to inefficient processes. An array of factors can impact the overall use of the system. In regards to the operator, their attitudes about the automation can determine how they interact with the system during the task as well as over time.

Cognitions

Several cognitive constructs are identified that can impact how an operator effectively handles tasks working with an automation, including situation awareness, mental workload, and the potential threat of skill decay. In the cases of monitoring tasks, multitasking, and in many other contexts, situation awareness is a cognitive construct considered critical in operators' interaction with automation systems. Endsley (1995a) theorized that situation awareness is composed of 3 different stages: (1) perception (the operator's ability to identify the status of elements in the task environment), (2) projection (comprehending the perceived elements into meaningfull information), and (3) prediction (ability to determine future states of the entire system with currently obtained information).

One challenge associated with situation awareness is obtaining valid real-time measuresments instead of self-report, as post-task questionnaires will only elicit responses of what operators were aware of, and may not indicate what information they have overlooked (Endsley, 1995b). More task-oriented approaches like the Situation Awareness Global Assessment Technique (SAGAT), pauses the simulated task and solicits inquiries about the situation for the operator to answer. While this approach may be considered more robust in assessing their SA, limitations can include the test as being a distractor from the task due to interruptions of information not needed during the task process (Endsley, 1995b).

Workload is an important consideration for automation, as technological systems are able to both benefit and hinder the operator's workload based on a variety of factors. These factors include task allocation and system interface design (Parasuraman, Sheridan, & Wickens, 2008). As it may be a challenge to measure workload objectively, a common approach is to rely on subjective workload metrics such as the NASA TLX (Hart & Staveland, 1988) to do so. However, physiological measures are available that can measure indicators of mental workload in real time (Veltman & Gaillard, 1996).

In cases where an automation system replaces tasks originally allocated to a human operator, the threat of skill decay is an important factor influencing safety and efficiency of a system from a long-term perspective. This can be an important threat in cases where the operator transitions to the role of monitoring a system that conducts a task they originally conducted manually. Skill decay is defined as a decrease or loss of trained skills and knowledge due to prolonged periods of nonuse (Arthur & Bennett, 1998). Kaber and Endsley (1997) discussed that decreased opportunities of direct/manual controlling of the system can have an impact on efficient operation and failure recovery. One of the direct causes of decreased manual control is associated with the design of the system, specifically the level of automation (Parasuraman, 2000). This in turn will negatively impact the entire safety of the system and potential failure rates.

Arthur & Bennett (1998) discovered multiple factors that can influence skill decay based on the task type. First, closed-loop tasks (i.e., tasks that have a definite beginning and end) exhibited more skill retention than open-looped tasks (i.e., tasks that are less finite or not a clear solution), although the authors note that this conflicts with previous research findings, and this result may be due to an interaction between multiple moderating variables. Additionally, cognitive tasks were found to be more susceptible to skill decay than physical tasks, and accuracy-based tasks were found to be three more times susceptible to skill decay than speed-based tasks.

OUTCOMES

Automation safety refers to the entire system's sustainability to maintain effective operations from failures and accidents (Mustafiz et al., 2008), and can be affected by the various inputs and processes associated with the human-machine system. Efficiency in this case refers to the ability of the entire human-automation system to complete the given task in a minimal amount of time or effort. Measurement of these outcomes can come in a variety of forms, including time-based metrics (i.e., how fast a system can complete a given task), error-based metrics (i.e., how many errors were made in a session), coverage metrics (i.e., what proportion of the overall goal has been achieved), and many others (Olsen & Goodrich, 2003). Accurate and valid measurements of these outcome varaibles are important to consider, especially in cases where safety and performance risks can have severe consequences in the real world work environment.

DISCUSSION

As advanced technology systems are being applied in more domains and workplaces, the issue of automation risks are important to consider for maintaining efficiency and safety. The need for effective metrics to assess the safety and efficiency of automation systems can be useful in the design and implementation of system in various work environments.

Future efforts associated with this project involve identifying the relationship between specific factors mentioned in this paper, and developing a metrics toolkit for optimizing the conditions for any given human-automation system. The toolkit is conceptualized to cover multiple stages of the system development life cycle, including design phases, prototype testing, and field implementation. Controlled testing with a developed or analogue system with effective robust measures will help in determining changes and improvements to the overall system before implementation.

Our overall goal is to develop a toolkit that will allow designers to create functional specifications for systems that optimize human-automation safety, and efficiency. This toolkit will attempt to provide a consolidated guide that will enhance the system processes during implementation in the give task environment.

ACKNOWLEDGEMENT

The views expressed in this work are those of the authors. This work was supported by funding from NASA Grant #NNX13AO51G.

REFERENCES

American Psychological Association. (2014). *Personality*. [online] Retrieved from: http://apa.org/topics/personality/

Arthur Jr., W., & Bennett Jr., W. (1998). Factors that influence skill decay and retention. *Human Performance*, *11*(1), 57-101.

Burke, C.S., Lum, H.C., Scielzo, S.A., Smith-Jentsch, K., & Salas, E. (2009). Examining measures of team cognition in virtual teams: A heuristic and guidelines. In D. Schmorrow, J. Cohn, & D. Nicholson, (Eds.), *The PSI Handbook of Virtual Environments for Training and Education: Developments fo the Military and Beyond.* (266-283). Westport, CT: Praeger Security International.

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*(6), 962-968.

Donmez, B., Pina, P.E., & Cummings, M.L. (2008). Evaluation criteria for human-automation performance metrics. In R. Madhavan, E. Tunstel, & E. Messina (Eds.), *Performance evaluation and benchmarking of intelligent systems* (21-40). Boston, MA: Springer.

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*(6), 697-718.

Endsley, M.R. (1995a). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32-64.

Endsley, M.R. (1995b). Measurement of situation awareness in dynamic systems. *Human Factors*, *37*, 65-84.

Endsley, M.R., & Kaber, D.B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, *42*(3), 462-492.

Ericsson, K.A., & Lehmann, A.C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, *47*, 273-305.

Fiore, S.M., Rosen, M.A., Smith-Jentsch, K.A., & Salas, E. (2010). Toward an understanding of macrocognition in teams: Predicting processes in complex collaborative contexts. *Human Factors*, *52*(2), 203-244.

Geffroy, J. & Motet, G. (2002). *Design of dependable computing systems*. Dordrecht: Kluwer Academic Publishers.

Goldberg, L.R. (2001). International personality item pool: A scientifically collaborator for the development of advanced measures of personality traits and other individual differences. Retrieved from http://ipip.ori.org/

Goodrich, M.A., & Olsen Jr., D.R. (2003). Seven principles of efficient human robot interaction. *IEEE International Conference on Systems, Man and Cybernetics*, *4*, 3942-3948.

Huang, F. & Hwang, S. (2009). Experimental studies of computerized procedures and team size in nuclear power plant operations. *Nuclear Engineering and Design, 239,* 373-380.

Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y.C., de Visser, E.J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, *53*, 517-527.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139-183.

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*(3), 126-131.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(1), 50-80.

Matthews, G., Deary, I.J., & Whiteman, M.C. (2003). *Personality Traits*. Cambridge, United Kingdom: Cambridge University Press.

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

Metzger, U., & Parasuraman, R. (2001). Automation-related "complacency": Theory, empirical data, and design implications. In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting.*

Mustafiz, S., Sun, X., Kienzle, J. & Vangheluwe, H. (2008). Model-driven assessment of system dependability. *Software & Systems Modeling*, *7* (4), 487-502.

Olsen, D.R. & Goodrich, M.A. (2003). Metrics for evaluating human-robot interactions. In *Proceedings of the Workshop on Performance Metrics for Intelligent Systems*. Gaithersburg, MA.

Oakley, B., Mouloua, M., & Hancock, P.A. (2003). Effects of automation reliability on human monitoring performance. *Proceedings of the Human Factors and Ergonomic Society, 47*, 188-190.

Parasuraman, R. (2000). Designing automation for human use: empirical studies and quantitative models. *Ergonomics*, *43*(7), 931-951.

Parasuraman, R. & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *52*(3), 381-410.

Parasuraman, R., Mouloua, M., & Hilburn, B. (1999). Adaptive aiding and adaptive task allocation enhance human-machine interaction. *Automation technology and human performance: Current research and trends*, 119-123.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *38*(4), 665-679.

Parasuraman, R. & Riley, V. (1997). Humans and automation: use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230-253.

Parasuraman, R., Sheridan, T. B. & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 30* (3), 286-297.

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

Rice, S., & Keller, (2009). Automation reliance under time pressure. *Cognitive Technology*, *14*(1), 36-44.

Rose, C.L., Murphy, L.B., Byard, L., & Nikzad, K. (2002). The role of big five personality factors in vigilance performance and workload. *European Journal of Personality*, *16*, 185-200.

Sheridan, T. & Parasuraman, R. (2005). Human-automation interaction. *Reviews of human factors and ergonomics*, *1*(1), 89-129.

Steinfeld, A., Fong, T., Keber,D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006). Common metrics for human-robot interaction.

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*(2), 71-96.

Veltman, J.A., & Gaillard, A.W.K. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology*, *42*(5), 323-342.

Vidulich, M. A., & Wickens, C. D. (1986). Causes of dissociation between subjective workload measures and performance: Caveats for the use of subjective assessments. *Applied Ergonomics, 17* (4), 291–296.

Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159-177.

Wiegmann, D.A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. Theoretical Issues of Ergonomic Science, *2*(4), 352-367.

Figure 1. Theoretical framework for measuring human-automation safety and efficiency.