Expanding Human Response to Automation Failures to Sociotechnical Systems

Nancy J. Cooke

Human Systems Engineering

Arizona State University

Abstract

Skraaning and Jamieson (2023) raise some interesting issues related to the response of humans to automation failures and offer a taxonomy of failure types that broadens its definition. In this commentary a further attempt to broaden the scope of automation failures is made that places failures within a sociotechnical system of multiple humans and multiple machine components including automation. A suggestion of how one might understand the system's response to automation failures is offered and the inclusion of autonomy is raised as another complication.

Running head: Expanding Automaton Failures to Sociotechnical Systems

Contact: Nancy J. Cooke, ncooke@asu.edu

Expanding Human Response to Automation Failures to Sociotechnical Systems

Skraaning and Jamieson (2023) examine the concept of automation failures and human performance in response to those failures. Their claim is that automation failure as defined as a failure of solely the support system is overly narrow. Also, they argue that automation failure defined as any human misconception about automation is too broad. They go on to provide examples of automation failures in the aviation domain in which the automation worked as designed but was provided inaccurate data by another failed part of the system. These they labeled as “Systemic Automation Failures,” a term that I fully endorse. Skraaning and Jamieson (2023) go on to provide a taxonomy of automation failures that include the Systemic Automation Failures in addition to Elementary Automation Failures and failures arising from Human-Automation Interaction Breakdowns. It is interesting that “Human and Organizational Slips” are excluded from the taxonomy of automation induced challenges - more on this later.

From Human Automation Interaction to Human Automation Systems

I agree with Skraaning and Jamieson (2023), on the need to broaden our concept of automation failure and human performance challenges but would go at least one step further by proposing that most automation that humans interact with today and that presents the majority of challenges to human performance is embedded in a large and complex sociotechnical system (Carayon, 2006). Automation is not monolithic but occurs within a system of other machine components and other automation. In most cases, including the ones referenced by Skraaning and Jamieson (2023) – nuclear power plants and aviation – the system is one consisting of not only multiple machines and automated components, but also multiple humans, often with complex work interdependencies. In these sociotechnical systems it is difficult to parse the single operator interacting with a single automation out of the entire system. As an alternative to predicting the human's response to automation failure, we might predict the system's response to Systemic Automation Failures.

This systems thinking has led to a recent focus on human-automation teaming rather than human – automation interaction (O'Neill, McNeese, Baron, & Shelbe, 2022). A team is a system with team members that have a common goal and with varying degrees of interdependency (Cooke, Cohen, et al., 2022). Taking a systems perspective, it follows that teams need to be measured at the team level in the context of their task (Cooke, Gorman, et al., 2013). One way to measure at the team level is to focus that measurement on system interactions such as team communication or joint activity (Cooke & Gorman, 2009).

Indeed, it is because of complex and often unanticipated system interactions that it is difficult to isolate system components and point to a single point of failure. When perturbations (or failures) are introduced to the system, it is the system that responds to the failure by either adapting by reorganization and overcoming the failure or breaking down. Often a single failure will cascade through a system impacting other components along the way. This is depicted in Reason's (1990) "Swiss Cheese Model" of accidents in which accidents occur through multiple holes or weaknesses in layers of defenses that happen to align and result in an accident. This is realized in most aviation, medical, and industrial accidents as reported in Casey's (1998) stories in his book, "Set Phasers on Stun." The aviation examples of Skraaning and Jamieson (2023) are not single failure points. The Boeing 737 MAX accident that they describe resulted from a failed sensor and the failure of Boeing to draw attention to the new automation and its expected behavior (Skraaning & Jamieson, 2023).

Broadening the view of automation failures from human automation interaction to human-automation systems may enable the reconsideration of the "Human and Organizational Slips" that Skraaning and Jamieson (2023) excluded from the taxonomy of automation induced challenges. That is human slips due to poor mental models of automation may be driven by the design or affordances of the automation (e.g.., Tesla's "auto pilot") design as well as training that does not map well onto the automation (e.g., Boeing 737 MAX) or that cannot keep up with design changes (e.g., Tesla updates). These human slips may not be unrelated to features of the automation.

Predicting System Response in Sociotechnical Systems

Given the complexity of system failures and the system's responses to failures in sociotechnical systems, how might system response to failures be studied? Gorman, et al. (2019) have developed an approach to examining cascading failures through layers of a system – a layered dynamics approach. One example of the application of this approach to a sociotechnical system is in a recent AFOSR-supported project that examined distributed space operations in the face of failures or perturbations (Yin, et al., 2022). This system is distributed in time (i.e., varied communication latencies) and space and includes automation (robots) and multiple humans. The scenario, informed through interviews with space operation experts, included the following human and machine components, mostly played by humans: NASA Mission Control, Jet Propulsion Lab (JPL), International Space Station with two astronauts (one a space walker), a Lunar colony with one human and one untrustworthy robot, a Lunar Orbiter, a Mars Rover (played by a Husky Robot controlled with built-in latencies by JPL), and a Mars Orbiter. Each player had a script of actions and communications that took place over custom push-to-talk radios. Perturbations (e.g., asteroid strike on moon, untethered space walker) were introduced at set times in the scenario. Measures were taken continuously of communications (i.e., who was talking to whom), vehicle positioning (for the orbiters and Rover), and heart rate variability for the untethered space walker and human who is rebuilding equipment on the lunar surface to quickly restore oxygen supply after asteroid strike.

The layered dynamics approach involves examining system reorganization or adaptation in the face of a perturbation. Reorganization of the system is measured in terms of significant changes in entropy across communication, positioning, and physiological signals. Figure 1 depicts the communication channel usage over time (raw data on who is talking to whom). The series of communication channel patterns serves as a signal for the communication layer. Significant changes in entropy of that signal are taken as points of reorganization – often in response to a perturbation (Figure 2). With this analysis applied to all the system layers it is possible to see how communication, positioning, and physiology are impacted by a system failure and exactly which components of the system are impacted. Cascading impact can also be seen in changes in entropy in regard to the timing of the perturbation. Future work in this area will apply machine learning techniques to detect and predict system anomalies so that they can be quickly attended to and mitigated by mission command.

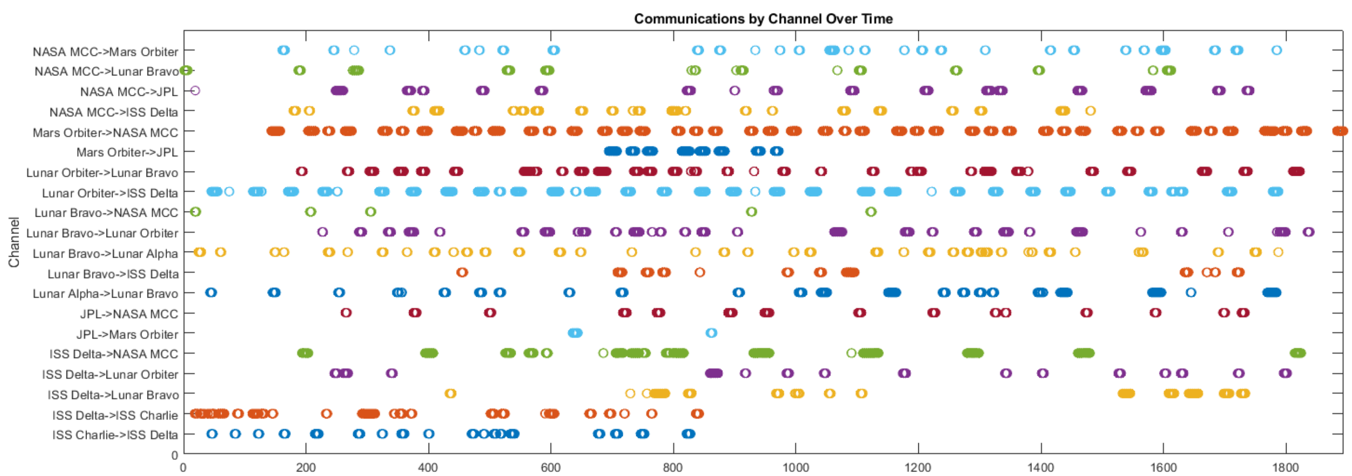

Figure 1. Patterns of communication channels over time.

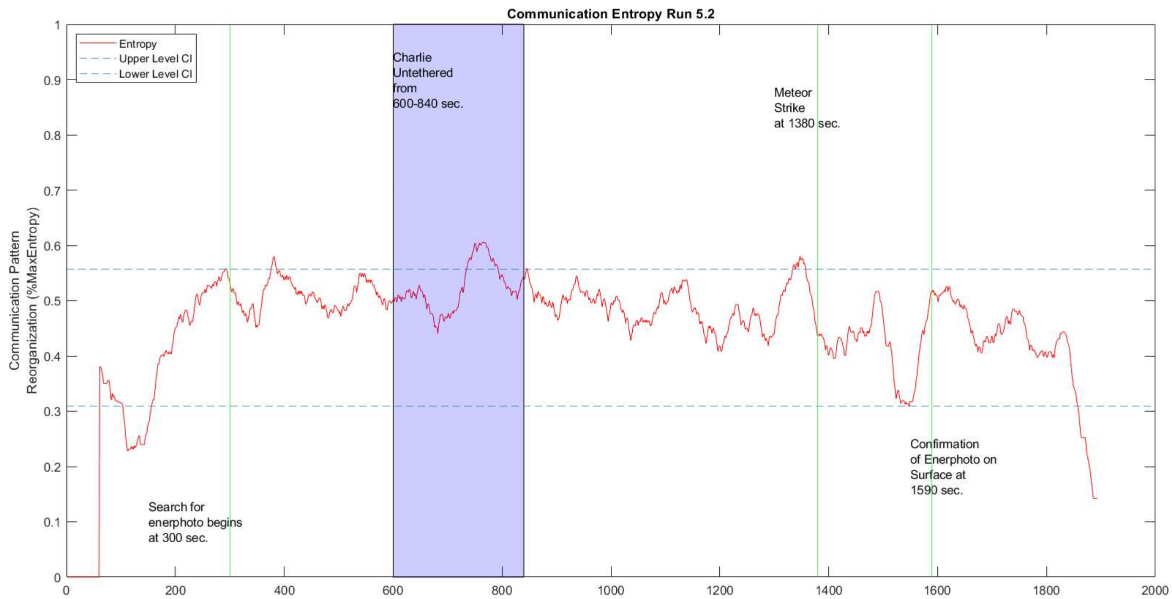

Figure 2. Examining changes in entropy of the communication patterns over time communication layer.

Automation vs. Autonomy

One more issue that complicates this a bit is the increasing talk of autonomy, the difference between automation and autonomy, and what role they each play in sociotechnical systems. There is no fully autonomous system at this time. One could argue that humans are not fully autonomous in the sense that they depend on others in their social system as well as technology to survive. Systems like the Mars Rover have some functions that are carried out autonomously and some that are fully controlled by humans. There have been several proposed levels of automation and autonomy over the years (Wickes, et al., 2010; Vagia, et al., 2016). In most systems there are functions with mixed levels of autonomy. Autonomy is generally considered to have more intelligence than automation. That said, what happens when intelligent autonomy fails? Are human expectations of autonomy different than automation? Is the response different?

In an experiment by Demir, et al. (2021) in simulated remotely piloted aircraft System (RPAS) ground control in which a three-member teame took reconnaissance photos from a single RPAS, one team member was replaced by a confederate who acted as autonomy and failed in specific ways (e.g., comprehension errors). In other cases, aspects of the display that provided location information failed (i.e., automation failure). Both autonomy and automation failures served as perturbations for the participants. However, the two human team members treated the failures differently. The automation failures were handled by adaptation of coordination procedures. They would work around failure by communicating information to each other that they had and that was needed. However, in cases of autonomy failures, the only way teams overcame the failure was to persist in badgering the autonomy to carry out the requested action. However, in many cases the teams gave up and moved on to the next target, failing to get a photo. In some cases the human participants indicated that they must be at fault, as the artificial intelligence that was making the mistakes should know better than they about this task as they were new to it. The expectations for autonomy exceeded the expectations for automation and thus the response to failure was very different.

Conclusion

Skraaning and Jamieson (2023) certainly raise valid points about the need to expand the definition of automation failures. In their introduction of the term Systemic Automation Failures, they hint at the complexity and interdependencies of systems and relevance to defining automation failures. They also separate human failures from the system. In contrast, looking at automation (or autonomy) failures through a sociotechnical lens reveals possibilities for other ways to study these systems and to predict system response. Finally, intelligent and more autonomous systems open up new possibilities for human and system responses due to changing expectations. Skraaning and Jamieson (2023) make a very interesting case for broadening our view of automation failures and human response. There is even more that we should consider in today's complex systems of humans and technology.

References

Carayon, P. (2006). Human factors of complex sociotechnical systems. *Applied ergonomics*, *37*(4), 525-535.

Casey, S. (1998). *Set Phasers on Stun and Other True Tales of Design, Technology, and Human Error.* Santa Barbara, CA: Aegean Publishing Co.

Cooke, N. J., Cohen, M.C., Fazio, W.C., Inderberg, L. H., Johnson C. J., Lematta, G. J., Peel, M., Teo, A. From Teams to Teamness: Future Directions in the Science of Team Cognition. (2022). *At the Forefront of Human Factors/Ergonomics, Human Factors.*

Cooke, N. J. & Gorman, J. C. (2009). Interaction-Based Measures of Cognitive Systems. *Journal of Cognitive Engineering and Decision Making: Special Section on: Integrating Cognitive Engineering in the Systems Engineering Process: Opportunities, Challenges and Emerging Approaches 3, 27-46.*

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J.L. (2013). Interactive Team Cognition, *Cognitive Science, 37, 255-285,* DOI: 10.1111/cogs.12009.

Demir, M., McNeese, N. J., Gorman, J., Cooke, N. J., Myers, C. & Grimm, D. (2021). Exploration of team trust and interaction dynamics in human-autonomy teaming. *IEEE*

Transactions on Human-Machine Systems, 51, 696-705. Doi: 10.1109/THMS.2021.3115058

Gorman, J. C., Demir, M., Cooke, N. J., & Grimm, D. A. (2019). Evaluating Sociotechnical Dynamics in a Simulated Remotely-Piloted Aircraft System: A Layered Dynamics Approach. *Ergonomics*, *0*(ja), 1–44. https:// doi.org/10.1080/00140139.2018.1557750

O'Neill, T., McNeese, N., Barron, A., & Schelble, B. (2022). Human–autonomy teaming: A review and analysis of the empirical literature. *Human factors*, *64*(5), 904-938.

Reason, James (1990). *Human Error*. New York, N.Y.: Cambridge University Press. ISBN 978-0-521-30669-0.

Skraaning, G., & Jamieson, G. A. (2023) The Failure to Grasp Automation Failure. Journal of Cognitive Engineering and Decision Making, pp 1-12.

Vagia, M., Transeth, A. A., & Fjerdingen, S. A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed?. *Applied ergonomics*, *53*, 190-202.

Wickens, C. D., Li, H., Santamaria, A., Sebok, A., & Sarter, N. B. (2010, September). Stages and levels of automation: An integrated meta-analysis. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 54, No. 4, pp. 389-393). Sage CA: Los Angeles, CA: Sage Publications.

Yin, X., Clark, J., Johnson, C. J., Grimm, D. A., Zhou, S., Wong, M., Cauffman, S., Demir, M., Cooke, N. J., Gorman, J. C. (2022). Development of a distributed teaming scenario for future space operations. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 66, No. 1). Sage CA: Los Angeles, CA: SAGE Publications.

Author Bio

Nancy J. Cooke is a professor of Human Systems Engineering at Arizona State University and is senior scientist of ASU's Center for Human, Artificial Intelligence, and Robot Teaming. Dr. Cooke studies individual and team cognition and its application to human, AI, and robot teaming and conducts empirical assessments of teams and teamwork.