The 2016 AAAI Fall Symposium Series:
Cross-Disciplinary Challenges for Autonomous Systems
Technical Report FS-16-03

Seeking Human-Centered Autonomous Systems Capabilities in a Machine-Centered Development Environment

Shu-Chieh Wu
San Jose State University and NASA Ames Research Center
shu-chieh.wu@nasa.gov

Abstract

This paper aims to shed light on the cross-disciplinary challenges involved in the development of autonomous systems from a practice standpoint. To that end, the paper examines what aspects of human-centered autonomous systems capabilities may be difficult to achieve using a machine-centered development process, the common practice. The paper concludes with suggestions for what more may be done to enable human-centered design considerations to be more effectively infused in the development process.

Introduction

Automated or autonomous systems are increasingly being used to facilitate or perform operations that have previously been done by human operators in complex aerospace environments. To achieve optimal performance in joint human-machine systems, it is recognized that humans and machines need to be able to collaborate like a team. Approaches to designing autonomous systems that will support team-like collaboration with humans, such as adaptive supervisory control, adaptive automation, dynamic task allocation, adjustable autonomy, and mixed-initiative interaction (National Research Council 2014), place emphases sensibly on making the machines adapt to human constraints and their situational and momentary task needs. However, the development of such systems often follows a machine-centered path, with the autonomous capabilities developed before considering how they may be used effectively by human operators.

The goal of this paper is to examine the implication of using a machine-centered approach to developing human-centered autonomous systems capabilities. Specifically, the paper seeks insights to two questions:

1) What, if any, aspects of human-centered autonomous systems capabilities may be difficult to achieve when human-machine collaboration issues are handled after the capabilities have been developed?

2) Given the predominantly machine-centered development environment, how may human-centered design considerations be more effectively infused in the development processes?

To seek insights to the first question, the paper uses adaptive automation, a widely adopted teaming approach, as the illustrating case and reviews its nature and requirements, with the goal of identifying when in the development process the requirements need to be incorporated to achieve the intended qualities in the final system. To seek insights to the second question, the paper considers what more can be done from the human factors side in terms of requirement specifications and from the systems engineering side in terms of incorporating human operators into system considerations.

It should be noted that the reality of machine-centered development is neither new nor unique to the development of autonomous capabilities in complex systems. Despite its relevance, human factors has long played a diminished, if not entirely absent, role in systems design stages. Feigh et al. (2011) note that the ineffective infusion of human factors into systems engineering is attributable in part to a view widely held by systems designers that humans are external to the systems being developed. Holding such a view is understandable at a time when machines performed functions not typically done or doable by humans. However, as machines are now being designed to take over operations, some performed previously only by humans, the classic operational boundary between humans and machines no longer applies, thus making better integration between human factors and complex systems design a renewed priority and challenge.

Attributes and Requirements for Adaptive Automation

As the word adaptive suggests, adaptive automation is flexible or dynamic in nature, modifiable during operation according to changing situational demands (Scerbo 1996). Adaptation through dynamic task or function allocation between human operators and automation may be the most widely adopted form (e.g., Feigh and Pritchett 2014; Kaber et al. 2001; Raja Parasuraman and Wickens 2008) though variations exist that include adaptable supervisory control (a human operator determines the level of control delegated to automation; Miller and Parasuraman 2007) and similarly adjustable autonomy (a human operator adjusts the level of autonomy between manual and autonomous operations; Dorais et al. 1999). In the following, adaptation through function allocation will be used as the basis for discussion.

The design of adaptive automation requires careful considerations over at least two aspects. One aspect concerns the attributes (Rouse and Rouse 1983) or strategies (Scerbo 2007) of adaptation, which may be thought of as the rules that govern the modifications underlying the adaptation. For adaptive function allocation, one of the first things to be decided is ***how is the adaptation done?*** Methods for adaptively adjusting system functionality responsibilities among agents (human and autonomous system) in a joint system have included task allocation (allocating an entire task to the human operator or the autonomous system) and task partitioning (partitioning and distributing portions of a task to the operator or system). A task could also be transformed to make it easier for the operator to perform, or more challenging so to re-engage the operators (Steinhauser, Pavlas, and Hancock 2009). Certain basic assumptions need to hold true regardless of the nature of the adjustments. According to Feigh and Pritchett (2014), each agent must be able to perform the individual as well as the collective functions assigned. All agents must also be able to handle the coordination incurred as a result of functional allocation. To choose the best suited method among the alternatives would require at a minimum a good knowledge of what each agent is capable of under the optimal condition, what each agent is capable of under a specific condition (which may be influenced by performing its assigned function), and how the agents perform their respective functions as a team. It is also necessary to have a clear understanding of the task itself to devise sensible distribution or partitioning schemes.

Second, ***what triggers the adaptation?*** According to Scerbo (2007), adaptation could be triggered by external events (e.g., certain tasks or emergency situations) or by changes in predicted operator performance. The latter implies there needs to be a human performance model for predicting how human operators will perform based on the state of task demands and the availability of information processing resources for the human operator, as well as online assessments of what the human operator is doing, how he or she is doing (e.g., workload), and what he or she intends to do next (Rouse and Rouse 1983). Understandably, to decide among the triggering alternatives would first require knowing how the adaptation is to be done as well as theories and methods for assessing human performance.

Third, ***who initiates the adaptation?*** Even though the triggering conditions discussed above suggest that the system itself could in theory autonomously initiate the adaptation, opinions vary on whether it should (Scerbo 1996). Some argue that the operator should always be in control to avoid the types of negative consequences associated with being out of the loop (e.g., Endsley and Kiris 1995; Sarter and Woods 1995) whereas others found that managing functional allocation while performing primary task responsibilities could in fact increase operator workload (Kaber and Riley 1999). Once again, to decide among the alternatives would likely require first knowing how the adaptation is to be done and to be triggered, in addition to having the ability (and/or capability) to assess whether and how the proposed adaptation scheme would impact the human operator's performance.

Lastly, even for non-adaptive autonomous capabilities, basic requirements are needed for determining what the human operator's task is and specifying ***what aspects of the task the autonomous systems capabilities are designed to support and how they support them***. For example, decision aiding could take a variety of forms (alert, acquisition, evaluation, synthesis, advising) differing in the level of analysis required (Rouse and Rouse 1983).

Conceptually orthogonal to the attributes or strategies of adaptation is the aspect that concerns implementing the attributes or strategies in ways that support collaboration in joint human-machine systems. Klein, Woods, Bradshaw, Hoffman, and Feltovich (2004) compare taking part in a joint system as entering into a basic compact, one founded on an agreement to "facilitate coordination, work toward shared goals, and prevent breakdowns in team coordination" (Klein et al. 2004, p.91). Understandably, the ability for the humans and machines in the system to uphold the agreement depends critically on whether the system environment possesses the necessary qualities to facilitate coordination. According to Klein et al., these qualities are ***common ground (or interpretability), mutual predictability, and directability***. The system environment needs to provide necessary information and feedback so that the parties involved could establish and maintain common ground, which will in turn enable each party to properly interpret the other's actions and detect anomalies. The system environment also needs to provide consistent interactions through experience so that the parties involved could reasonably predict the other's action and in turn plan own actions. Last but not least, the system environment needs to allow the parties involved to be able to modify the other's actions and conversely be responsive to the other's requests.

Given the properties Klein et al. (2004) proposed, what needs to be known during the formulation of the requirements so that the final system will support collaboration? In their coactive design approach proposed to address the very interdependencies in collaboration, Johnson et al. (2014) list three items related to determining the requirements of the interdependencies: 1) Identify what each party needs to make observable to the others; 2) Identify behaviors that need to be predictable by the others; and 3) Identify the ways parties need to direct each other. Understandably, the answer to each of these items will at a minimum depend on determinations made regarding the attributes/strategies of adaptation. It will also depend on one party's knowledge and expectation for another (i.e., common ground) that have obtained through interactions. For example, if a portion of the task is to be assigned to the automation when the system detects that the operator has neglected it for some period of time based on certain behavioral measures, the system will need to make clear to the operator what it has observed of the operator (supporting interpretability), its intention to take over the neglected task (supporting directability) and, if permission granted, when the handoff occurs (supporting predictability).

Design Implementation Considerations

It should come as no surprise that non-adaptive autonomous systems capabilities could be developed, as they often are, with only an understanding of the task performed by the human operator. Also not surprisingly, almost all adaptive attributes will likely be difficult to achieve if they are not considered prior to the development of the autonomous systems capabilities. The reasons for the difficulty are somewhat nuanced, though. First, there is a web of interdependencies among the attributes. It may appear that task adaptation is at the core as it affects the choice of the triggering condition and the initiator. However, as choices implemented to realize adaptive automation sometimes become additional tasks, they could in turn impact the chosen scheme of task adaptation. As a result, the relations among the attributes are as dynamic as the system behavior they are tasked to create.

Second, the choice on task adaptation is made based in part on the known capabilities of the agents involved. The capability of the autonomous system could certainly change with development, so may the capability of the human operator due to interacting with the system. These changes may again propagate to influence the choices and behaviors of the implementations of the other attributes.

As to whether the collaboration support qualities can be achieved after the autonomous systems capabilities have been developed, the answer is probably both yes and no. On the one hand, there is likely always room for improvements in system interface design to support better operator situation awareness and communication of system states and task status (Kaber et al. 2001). On the other hand, if the autonomous capabilities are not developed based on a needs analysis of the operator task, even with improved interface design the final system may be at best human-user-friendly and not human-centered.

This paper uses adaptive function allocation as the illustrating example for human-centered automation. Function allocation as a proposal for human-machine collaboration has faced a fair number of criticisms, mostly related to its potentially over-simplified assumption that functions can be cleanly decomposed and assigned to different agents (Bradshaw et al. 2013; Dekker and Woods 2002). The validity of the criticisms notwithstanding, they are not particular pertinent to the discussions here. If anything, the simplified assumption may have facilitated getting the points across by reducing the complexity in describing possible human-machine collaboration schemes.

Adaptation: The Answer to More than One Question

As previously mentioned, the machine-centered development practice has had a long history, arising from realities that are as much institutional as ideological. Certainly the practice cannot be changed overnight by one paper. Having said that, the following considers what more may be done to more effectively infuse human-centered design considerations in the development process from both the human factors side and the systems engineering side. Once again, adaptation may be the answer.

From the human factors side, the efficiency in infusing human-centered design considerations may be improved if human factors practitioners are also well versed in the engineering process. The potential benefits are twofold. On the one hand, having a good knowledge of the engineering process could help formulate requirements in ways that can be more readily incorporated by the system developers. On the other hand, the knowledge may also help establish realistic expectations of the range of capabilities in the final system given the requirements incorporated (or not), in a way similar to what the analysis here intended to illustrate. In today's machine-centered development environment for complex systems, the expectations of human factors have evolved from simply improving the user interface to improving the transparency of the user interface. Such expectations reflect negligence of the possibility that the cause of an opaque interface may have more to do with the underlying design of the autonomous systems capabilities rather than the interface per se. Having a good knowledge of how systems are developed based on requirements could potentially help envision the final system more correctly.

From the systems engineering side, the efficiency in infusing human-centered design considerations may be improved if human operators could be included as part of the system considerations. In an eloquently written position paper, Griffin (2010) contends that systems engineering as a discipline faces significant challenges in developing complex systems, evidenced by the failures in important and complex systems like the space shuttles where everything in the way of process control was thought to have been done correctly. He argues that systems engineering should be concerned "with context over structure, with interactions over elements, with the whole over the sum of the parts" (p.2). In other words, Griffin also thinks that systems engineering should adapt in order to meet the new demands of complex system development. Rather than simply satisfying well defined requirements and processes, Griffin sees a more holistic goal – achieving an elegant design. He considers a design to be elegant if it possesses four qualities: the design works, is robust, is efficient, and it accomplishes its intended purposes while minimizing unintended consequences. Griffin's (2010) aspiration for systems engineering, especially the goal of an elegant design, echoes the ideal image for the future autonomous systems. If humans are to remain as part of this elegant system, their interactions with the rest of the system must be considered.

Discussion

The pursuit for autonomous systems capabilities is seeing unparalleled determination and a rapid pace, thanks in part to the resurgent interest in artificial intelligence powered by cheap parallel computation, big data, and better machine learning algorithms (Kelly 2014). With time many envision that fully autonomous systems will deliver significantly improved safety, efficiency, and convenience in personal transportation, aviation, and many other areas of daily lives. Before then, however, greater danger lurks during the transition period when automation works most, but not all of the time. The issue, as Norman (2015) argues, lies not with automation itself but with the automation-centered development approach, which in the process of maximizing its use fails to consider how an imperfect automation combined with human cognitive limitations can lead to catastrophic consequences. Echoing the recommendations made by the United States Defense Science Board in their 2012 report (Defense Science Board 2012), Norman advocates for a human-centered development approach, one that seeks to optimize the total performance of human-machine collaboration whereby each does what it is best at.

The goal of the present paper is to shed light on the cross-disciplinary challenges involved in the development of human-centered autonomous systems capabilities, much like what Norman (2015) advocates, in this transition period from a practice standpoint. Machine-centered development practices are widely adopted and, in the case of aircraft automation, may possibly have contributed to many of the automation related accidents (see Billings 1996). Even though human-machine collaboration is well recognized as an important capability to support in complex systems, to my knowledge no research has considered if today's machine-centered development practice could realistically achieve human-centered autonomy in the final system. The present paper represents a preliminary attempt at that. Needless to say much more work is needed to more systematically identify the design requirements for human-centered autonomy, address the dependencies and interplays between them, map them on to the development process, and to finally use the understanding acquired to inform the design process.

Acknowledgement

The preparation of this paper was supported by the Human Systems Integration Division at NASA Ames Research Center. The author thanks Brent Beutter for comments on a previous version of the paper.

References

Billings, Charles E. 1996. *Aviation Automation: The Search for A Human-Centered Approach*. 1 edition. Mahwah, N.J: CRC Press.

Bradshaw, Jeffrey M.; Hoffman, Robert R.; Johnson, Matthew; and Woods, David D. 2013. The Seven Deadly Myths of "Autonomous Systems." *IEEE Intelligent Systems* 28 (3): 54–61. doi:10.1109/MIS.2013.70.

Defense Science Board. 2012. The Role of Autonomy in DoD Systems. Task Force Report. http://fas.org/irp/agency/dod/dsb/autonomy.pdf.

Dekker, S. W. A.; and Woods, D. D. 2002. MABA-MABA or Abracadabra? Progress on Human–Automation Co-Ordination. *Cognition, Technology & Work* 4 (4): 240–44. doi:10.1007/s101110200022.

Dorais, Gregory; Bonasso, R. Peter; Kortenkamp, David; Pell, Barney; and Schreckenghost, Debra. 1999. Adjustable Autonomy for Human-Centered Autonomous Systems. In *Working Notes of the Sixteenth International Joint Conference on Artificial Intelligence Workshop on Adjustable Autonomy Systems*, 16–35.

Endsley, Mica R.; and Kiris, Esin O. 1995. The Out-of-the-Loop Performance Problem and Level of Control in Automation.

Human Factors: The Journal of the Human Factors and Ergonomics Society 37 (2): 381–94. doi:10.1518/001872095779064555.

Feigh, Karen M.; Chua, Zarrin; Garg, Chaya; Jacobsen, Alan; O'Hara, John; Rogers, William; and Shutko, John. 2011. Current State of Human Factors in Systems Design. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 55 (1): 267–71. doi:10.1177/1071181311551055.

Feigh, Karen M.; and Pritchett, Amy R. 2014. Requirements for Effective Function Allocation A Critical Review. *Journal of Cognitive Engineering and Decision Making* 8 (1): 23–32. doi:10.1177/1555343413490945.

Griffin, Michael D. 2010. How Do We Fix System Engineering. In *61st Annual International Congress, Prague, Czech Republic*. Vol. 27. https://info.aiaa.org/tac/ETMG/CASE/Shared%20Documents/CASE%202012/Planning%20Documents/How%20Do%20We%20Fix%20System%20Engineering.pdf.

Johnson, Matthew; Bradshaw, Jeffrey M.; Feltovich, Paul J.; Jonker, Catholijn M.; Riemsdijk, M. Birna van; and Sierhuis, Maarten. 2014. Coactive Design: Designing Support for Interdependence in Joint Activity. *Journal of Human-Robot Interaction* 3 (1): 43–69. doi:10.5898/JHRI.3.1.Johnson.

Kaber, David B.; and Riley, Jennifer M. 1999. Adaptive Automation of a Dynamic Control Task Based on Secondary Task Workload Measurement. *International Journal of Cognitive Ergonomics* 3 (3): 169.

Kaber, David B.; Riley, Jennifer M.; Tan, Kheng-Wooi; and Endsley, Mica R. 2001. On the Design of Adaptive Automation for Complex Systems. *International Journal of Cognitive Ergonomics* 5 (1): 37–57.

Kelly, Kevin. 2014. The Three Breakthroughs That Have Finally Unleashed AI on the World. *Wired*, October 27. http://www.wired.com/2014/10/future-of-artificial-intelligence/.

Klein, G.; Woods, D.D.; Bradshaw, J.M.; Hoffman, R.R.; and Feltovich, P.J. 2004. Ten Challenges for Making Automation A "team Player" in Joint Human-Agent Activity. *IEEE Intelligent Systems* 19 (6): 91–95. doi:10.1109/MIS.2004.74.

Miller, Christopher A.; and Parasuraman, Raja. 2007. Designing for Flexible Interaction between Humans and Automation: Delegation Interfaces for Supervisory Control. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 49 (1): 57–75. doi:10.1518/001872007779598037.

National Research Council. 2014. *Complex Operational Decision Making in Networked Systems of Humans and Machines: A Multidisciplinary Approach*. Washington, D.C.: National Academies Press. http://www.nap.edu/catalog/18844.

Norman, Donald A. 2015. The Human Side of Automation. In *Road Vehicle Automation 2*, edited by Gereon Meyer and Sven Beiker, 73–79. Lecture Notes in Mobility. Springer International Publishing. http://link.springer.com/chapter/10.1007/978-3-319-19078-5_7.

Parasuraman, Raja; and Wickens, Christopher D. 2008. Humans: Still Vital After All These Years of Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 50 (3): 511–20. doi:10.1518/001872008X312198.

Rouse, William B.; and Rouse, Sandra H. 1983. A Framework for Research on Adaptive Decision Aids. Technical Report AFAMRL-TR-83-082. Ohio: Wright-Patterson Air Force Base. http://www.dtic.mil/dtic/tr/fulltext/u2/a138331.pdf.

Sarter, Nadine B.; and Woods, David D. 1995. How in the World Did We Ever Get into That Mode? Mode Error and Awareness in Supervisory Control. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37 (1): 5–19. doi:10.1518/001872095779049516.

Scerbo, Mark W. 1996. Theoretical Perspectives on Adaptive Automation. In *Automation and Human Performance: Theory and Applications*, edited by R. Parasuraman and M. Mouloua, 37–63. Human Factors in Transportation. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

———. 2007. Adaptive Automation. In *Neuroergonomics: The Brain at Work*, edited by Raja Parasuraman and Matthew Rizzo, 239–52. Human-Technology Interaction. Oxford, New York: Oxford University Press.

Steinhauser, Natalie B.; Pavlas, Davin; and Hancock, P. A. 2009. Design Principles for Adaptive Automation and Aiding. *Ergonomics in Design: The Quarterly of Human Factors Applications* 17 (2): 6–10. doi:10.1518/106480409X435943.