NASA/TM—2017–219481

Concepts for the Design of Human-Autonomy Systems

Mary M. Connors
NASA Ames Research Center

March 2017

NASA STI Program…in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include creating custom thesauri, building customized databases, and organizing and publishing research results.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov

- E-mail your question via the Internet to help@sti.nasa.gov

- Phone the NASA STI Help Desk at (757) 864-9658

- Write to:
  NASA STI Information Desk
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

NASA/TM—2017–219481

Concepts for the Design of Human-Autonomy Systems

Mary M. Connors
NASA Ames Research Center

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California

March 2017

Available from:

NASA STI Program
STI Support Services
Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199

This report is also available in electronic form at http://www.sti.nasa.gov
or http://ntrs.nasa.gov/

Table of Contents

Preface 1
1. Introduction 2
2. Planning for Design 2
    2.1 Common Features 2
    2.2 Models 3
    2.3 Metrics 3
    2.4 Risk 3
    2.5 Costs 3
3. Bench Design 4
    3.1 General Requirements 4
    3.2 Intent 4
    3.3 Information 4
    3.4 Transitioning 5
    3.5 Timing 5
    3.6 Monitoring 5
    3.7 Trust and Transparency 5
    3.8 Workload 6
    3.9 System Failure 6
    3.10 Validation and Safety 6
    3.11 Certification 6
4. Industry Design Considerations 6
5. Conclusions 7
References: Workshop Reports 8

Concepts for the Design of Human-Autonomy Systems

Mary M. Connors[1]

Preface

Over the last two years a number of workshops or similar meetings have been held under the auspice of NASA Ames Research Center that, although dealing with different primary foci, considered at some level the cross-cutting topic of humans-autonomous systems and the partnership between them. This publication presents findings from the reports of these meetings as they inform the process of designing autonomous systems to work in conjunction with humans. Some workshops were highly relevant to this question, others less so. Drawing from conclusions in these reports, this publication presents, at a high level, a summary of issues that could usefully be considered in the design process of automated systems working with human agents. Although differing in emphases, reports from the several workshops reflect many similar beliefs regarding human-autonomy teaming (HAT). Conclusions can be viewed as generally applicable to various aviation venues—air, ground, and air-ground interacions.

The objective of the workshops was primarily to identify issues, concerns, and research needs associated with evolving human-autonomy systems—not in providing solutions. The analysis herewithin is generally limited to the findings in the workshop reports. And, since there was often commonality in findings among the several reports, no attempt is made to ascribe particular findings or observations to particular reports.

The meeting reports consulted for this review are as follows:

- Transitioning to Autonomy: Changes in the Role of Humans in the Air Transportation System, April 2015.
- UTM Convention, July 2015.
- AIAA, ISTC, Intelligent Systems in Aerospace, January 2016.
- Assurance for Autonomy Systems, January 2016.

[1] NASA Ames Research Center, Moffett Field, California.

1. Introduction

Common to all the reports reviewed was the assumption that, although the capabilities of humans can be expected to remain more or less constant, autonomy and autonomous systems will continue to advance rapidly. Increasing the role of autonomy depends on the development of adaptive, non-deterministic systems. It is generally agreed that, although direct human involvement in these systems will be needed for years to come, this involvement will decrease over time, at least at the level of the operator. This predicted shift from human to autonomy changes the balance from the present environment where the human is assumed to compensate for safety gaps, to future operations where the automated system and its designers will bear increasing responsibility for both performance and safety. The path of the advancement of autonomy systems will depend to a large extent on advances in machine learning and computational intelligence tools. And, as today, the challenge for designers will be heavily weighted towards dealing with unexpected and hazardous conditions in dynamically changing environments. Further complicating these technological developments is the question of cyber-security, i.e., how to protect while developing.

In addition to effectiveness and safety, designs will also need to be reliable. The approach needed for future design goes well beyond the characteristics of individual components or of interconnected components and must include the full integration of hardware, software, and liveware into a highly functioning human-autonomty teaming system.

Combining human(s) and automation system(s) in a true teaming arrangement, at the level being discussed, is a new and challenging endeavor. It requires new architectures, new frameworks, new organizational categories, and new methods for standardization. While learning to operate in this emerging environment, designers should keep in mind that the more complex a system (in terms of goals, tasks, number and types of vehicles, environment and the like) the more difficult the implementation; conversely, the more closed the system (e.g., geofenced) and the more controllable the conditions of the system, the easier it is to implement. In other words, a guiding principle is to simplify to the extent possible for the system of interest.

2. Planning for Design

In considering the specifics of design, the potential contributions that could be made by both humans and machines need to be included. There is general agreement that, at least for the foreseeable future, adaptive automation should be human-driven. As such, the designer should consider the strengths and weaknesses, not only of the automation but also of the human partner. For instance, are the flexibility and resilience characteristics of the human important to the design? And, if so, how can these characteristics be preserved in the design? The designer should consider also the differences in how humans and automation maximize performance. Humans tend to sub-optimize, leaving themselves “wiggle room” to deal with unanticipated events. For autonomous agents, the choice tends to be to optimize at every level, from component to system. The designer needs to determine which approach is preferable for the particular design under consideration or whether there is a way to combine the two approaches.

2.1 Common Features

Planning design of a human-autonomy system should assure that, where applicable, the system is capable of moving easily from one generation to the next and should take into consideration if

and how the objectives or policies of the mission might change. Another early planning decision involves the extent to which the system will be prepared to address off-nominal, rare events. It is frequently opined that events that can be imagined should be considered in the design. But, while considering the possibilities, it may be impractical to plan for every considered event. So, the question remains as to whether events that can be imagined but are considered highly improbable should be included in the design.

2.2 Models

The reports reviewed addressed the need for new models (as well as new methods, simulation scenarios, and evaluation techniques) to support design progress and particularly to deal with probabilistic or uncertain events. Better models of humans, of autonomy, and particularly of human-autonomy systems are clearly needed to give designers a means of predicting how systems will be used and the consequences, both intended and untended, that are likely to follow.

2.3 Metrics

If automation and human-autonomy teaming are to advance, numerous improvements in metrics are needed. Evaluating the impact of non-deterministic systems presents important challenges to moving beyond the formal methods and analysis techniques presently available. For autonomy and for human-automation teams, performance-based results are needed that can quantify the benefits (or lack thereof) in autonomous and teamed systems. Specific to the design process, it is necessary to include an approach for determining if changes to autonomy are safe and beneficial to the system as a whole. This approach will be needed to assess efficiency and effectiveness as they affect total system performance in real-world environments.

2.4 Risk

Due to their greater probability of success, situations of low environmental risk offer more possibilities for automating than situations of higher risk. Developers can further lower risk by structuring conditions to be as narrow as feasible for the system, its intended application, and its potential future growth. And, development should avoid “feature creep” by basing development on the requirements of the system, not the capability or potential capability of computational intelligence. Field-testing will be required at different stages of development and field tests should be conducted in as low risk environment as appropriate.

2.5 Costs

Cost considerations should include all life cycle costs for both the machine and the human. Machine costs include maintenance, environmental (e.g., location fees, the need for wideband or special communications) and production resources; human costs include training for all agents involved. New methods of training are likely to be needed. Automation should be developed to minimize human training needs—initial, recurrent, and operational. More generally, planners should be able to quantitatively justify the benefits of advanced automation over traditional methods with reference to costs and mission requirements.

3. Bench Design

Every design begins with assumptions of what the system will be used for, what the environment will be like, and what the operator or controller of the system will be expected to do. It is important to make those assumptions explicit, and to assure that these assumptions are correct, especially with regard to human capabilities and what the human partner's responsibilities will be.

3.1 General Requirements

Unless otherwise determined, it can be assumed that the design is required to be resilient under uncertainty and able to capture and operate in complex and dynamically changing situations. While virtually all designs strive for efficiency, reliability, and scalability, other general requirements also need to be considered, especially in a human-autonomy teaming system. The autonomous system should:

- Provide for a shared mental model.
- Be able to convey its goals and have a way of knowing the goals of the human partner.
- Provide for intuitive ease of use (preferably involving natural language and aided by inputs from subject matter experts).
- Make clear who is in control when and how control is passed.
- Establish a balance between functionality and usability.
- Design-in sufficient involvement for the human team member to retain his/her proficiencies.
- Develop with knowledge of, and in parallel with, operational approaches.

3.2 Intent

A critically important and highly challenging requirement to human-autonomy teaming is the issue of conveying intent. Sharing intent is necessary to make actions predictable. Both agents must be able to discern what the other intends to do. From the standpoint of system design, the autonomy must be able to indicate to the human its intent and, to the extent possible, interpret the human partner's intent.

3.3 Information

A basic consideration in the design of a system is the organization and presentation of information, i.e., providing the right information to the right location at the right time. The information provided to the system must be accurate and timely and the system must be capable of interpreting and acting on it—not just pushing it (and the problem) out to the human. The machine system must have knowledge of both the activities to be accomplished and the environment in which the system will or could function. System designers should decide how much the human agent needs to understand of the machine's processes as well as how to support this understanding.

Both machine and human agents require the capability of sharing data for integrated analysis and decision-making. Both agents also require information concerning what the other is doing and when the other is reaching its limits.

Communication between agents and interconnectivity among sub-elements require some form of a common language, along with a substructure to support information requirements and exchange.

3.4 Transitioning

When activities are transitioned between agents or elements, this transition should be both observable and seamless. It is particularly important when the transition is system-to-human that the process be clear and, preferably, intuitive to the human.

3.5 Timing

The designer should be aware of the differences in how automation and humans deal with timing and scheduling processes. While humans can approach a solution from many sides, machines, presently and in general, require that processes be completed within certain time constraints and following prescribed sequences. Too often the timing constraints and sequencing requirements of the machine are not obvious to the human and are not made explicit by the system.

3.6 Monitoring

For some time to come, monitoring will be a primary means of assuring system safety. To this end, it is necessary that the human and the autonomous system be capable of monitoring each other. Autonomy, eventually, should be able to determine when the human is unable to function effectively. For instance, this might be done based on a psychophysical measure (e.g., reaction time) or by a physiological measures (e.g., eye movement). Human monitoring of the performance of the automated system presents the challenge of designing a level of monitoring that will not burden users nor bore them and cause them to lose attention. One suggestion in support of human monitoring is to "monitor by exception," i.e., to involve the human agent in monitoring the automation when something unusual happens. However, if the attention of the human is required only rarely, the system must be designed to bring the user back "into the loop," thus allowing the user to catch up on what is needed and to know what to do. Since automation tends to change rapidly, it has been suggested that implementation of advances in machine systems be linked to monitoring procedures so that the automation does not outpace the performance capability of the human monitor.

3.7 Trust and Transparency

It falls to the designer to ensure that the system is trustworthy, i.e., has a high level of validity and reliability. But future systems also will require a sufficiently high level of user "trust" in the automated actions being performed. An approach to encouraging trust in the system is for the machine partner to provide the user the bases for its decisions. Such transparency allows the human partner to know that what the automation is doing is (or is not) correct. Such transparency is needed both to establish and to maintain trust. While questions remain concerning how much transparency is needed or desirable to foster trust (without overloading the user) the design must provide the operator with sufficient information to know when the automation is making a questionable decision or when the automation itself is struggling with the decision. At a minimum, a process must be included for querying the automation as to the bases for and confidence in its actions. While encouraging trust, the system must actively engage the human to avoid "over-trusting" or complacency on the part of the human agent.

3.8 Workload

Design must take into consideration the workload imposed on the operator—both task workload and cognitive workload—remembering that teaming itself changes the nature of human work and can itself add to workload. But the workload considered should go beyond that of the operator and consider also any additional burden that might fall on, e.g., air traffic controllers or other service providers.

3.9 System Failure

Aeronautic systems should be designed to refuse-to-crash and to include emergency auto-land capability. One area of particular concern is the problem of lost links and data link failures. Another critical aspect associated with system failure is the ability to "fail gracefully." Failing gracefully implies a gradual deterioration process, but one that also keeps the human in the knowledge loop, allowing control to be transferred to the human partner and enabling uninterrupted system recovery.

3.10 Validation and Safety

The designer must provide for an active validation process for the system as a whole as well as for a crosschecking capability to ensure that safety-related items are not missed. Particular attention should be paid to autonomy-to-human and human-to-autonomy transitions. New concepts or elements should be field- or flight-tested in the full system as significant changes or additions to the autonomy element are made.

3.11 Certification

Certification will be a major barrier to adoption of automation in the aeronautics domain since no adaptive, non-deterministic system has yet been certified nor have the bases for certification of adaptive systems been established. Acknowledging the extreme difficulty in addressing this problem through existing procedures, a number of workshop participants suggested reversing the emphasis from specifying what a system will do to guaranteeing what a system will never do. A further suggestion involves licensing uses rather than certifying systems. Addressing the highly complicated certification issue goes well beyond system design and will involve the skilled attention of many levels of organization. However, the designer can at least consider, and attempt to address, what the concerns of the certifying (or licensing) authority will be.

4. Industry Design Considerations

Industry's approach to automation must take into account the need to transition from where we are now to where we will be as intelligent systems continue to advance. This long-term industry approach will encourage public acceptance and trust of intelligent systems generally. Although the designers of individual systems need to pursue simplicity (an emphasis on the particular), the industry as a whole should look for ways to develop general-purpose systems and to encourage the use of open architectures.

If automation is to reach its true potential, the industry must come together in two specific ways—standardization and sharing. In terms of standardization, a question that needs to be addressed is: How will multiple autonomous agents from different developers communicate with each other? Growth requires that information technology interfaces, taxonomies, and

lexicons be standardized[2]. The requirement for sharing presents a similar challenge. Developers will be learning-by-doing and a collaborative approach to sharing data, experiences and best practices is needed.

Certification, or a substitute practice, will constitute a daunting challenge for non-deterministic agents and for human-autonomy systems. Surmounting this barrier will be greatly aided by industry's willingness to formulate standards and to share experiences.

5. Conclusions

The reports of several workshops dealing with changes in the human-autonomy relationship were examined with a view towards design requirements. Some of these workshops were highly relevant to the question of human-autonomy teaming and its design; others were only tangentially related. But all had something to offer to this evolving area. From examining the reports of these workshops, it was concluded that results could be considered under three headings: (1) Planning for Design; (2) Bench Design; and (3) Industry Design Considerations. The information described provides a high-level assessment based on the workshop results. These findings can serve as pointers to areas that will require greater attention in the future than has been demanded in the past.

A general conclusion that was broadly endorsed by participants of the workshops is that testing of new and evolving systems should include representative field trials. A follow-on conclusion was that these trials should progress from less risky to more risky situations, from less challenging to more challenging operations. These conclusions are an acknowledgement of the high level of uncertainty associated with adaptive systems with reference to our present understanding of these systems and the complexity of real-world conditions in which these systems are expected to operate.

As automation advances, the responsibility for the safe performance of the human-autonomy team will move in the direction of the autonomous agent. While in the past the human operator has been assigned responsibility for filling gaps left by a deterministic agent, adaptive systems will now be expected to bear their share of the responsibility burden. Since only humans are ultimately accountable and liable, this re-scoping of roles represents a shift of responsibility from the user or operator to the developer or designer.

Since so much is yet to be learned, successful implementation of automation and human-automation teams depends to a large extent, on the ability of industry leaders to share information and experiences and to standardize where feasible. Such sharing and standardization are needed to begin to formulate a path to a certification solution.

[2] A serious complication to the goal of standardization is cyber-security. Where common elements with intelligent components are used, the opportunity for hacking, resulting in unsafe conditions, is increased and must be considered in the decision process.

References: Workshop Reports

Connors, Mary M. and Jon B. Holbrook, *Transitioning to Autonomy: Changes in the Role of Humans in the Air Transportation System*. NASA/TP—2015–218933, August 2015.

Report to the NASA Advisory Council–Aeronautics Committee, *UTM Convention*. July 28-30, 2015, NASA Ames Research Center, Moffett Field, CA.

American Institute of Aeronautics and Astronautics, Intelligent Systems Technical Committee, *Roadmap for intelligent systems in aerospace*. First Edition, Human-Systems Integration, June 6, 2016.

Brat, Guillaume, Misty Davis, Dimitro Giannakopoulou and Natasha Neogi. *Workshop on Assurance for Autonomous Systems for Aviation*. NASA/TM-2016-219446, May 2016.