Supporting Human-Automation Collaboration through Dynamic Function Allocation: The Case of Space Teleoperation

by

Huiyang Li

**A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in the University of Michigan
2013**

Doctoral Committee:

Professor Nadine B. Sarter, Chair
Professor Yili Liu
Associate Professor Bernard J. Martin
Professor Colleen M. Seifert

© Huiyang Li

2013

To Mom and Dad

Acknowledgements

This dissertation would not have been possible without the guidance, help and encouragement from some very important people and resources.

First and foremost, I would like to acknowledge my advisor and mentor, Dr. Nadine Sarter. Many thanks to her for taking me on as a student, for investing so much time and effort into my development as a researcher, and for her patience and faith in me throughout my Ph.D. journey. She not only sets high standards for me and all her students but enables us to do our best by providing careful and wise guidance, tireless support and help, detailed feedback, consistent challenges, timely encouragement, and necessary protection. I consider myself very fortunate to have had the opportunity to work with such an outstanding researcher and teacher and she will always be a role model for both my career and life.

I would also like to thank my committee members, Dr. Yili Liu, Dr. Bernard Martin, and Dr. Colleen Seifert for their time, support, and valuable input during the dissertation process and my graduate studies. They have not only contributed insightful comments and suggestions to my dissertation, but they have also provided very useful career guidance and support.

Outside of the University of Michigan, I have been fortunate to receive mentoring and training from Dr. Chris Wickens. It has been a privilege to be mentored by such a master in the field of human factors and engineering psychology. Chris has been an advisor whose input and support in my research and my career as a human factors scientist were invaluable. I would also like to thank Angie Sebok at Alion Science and Technology for her contribution to my dissertation work, and for being a great PI for our collaborative project.

I am also grateful to NASA Johnson Space Center, the IOE Department, and Rackham Graduate School for the financial support each provided to me during my graduate studies at the University of Michigan.

I also owe a debt of gratitude to the excellent staff in IOE, including Chuck Woolley and Eyvind Claxton who helped me design and build the research apparatus;

Olof (Mint) Minto who helped set up the specialized computer systems; Candy Ellis, Rick Kraus, Amy Warhaft who made sure that there was enough money for the project; and Matt, Tina, Liz, Gwen, Wanda, and Rod who worked hard to provide administrative support for research and teaching. Special thanks to Tina Blay for all the warmth she sheds through a smile, a greeting, and all the small things she does for everyone in the department.

I must also acknowledge the support I have received from our THInC Lab. I want to thank Joe Philips for being the most wonderful programmer and for his amazing ability to become an expert in a new system so quickly and to develop such impressive simulations. I would like to thank Nadine Moacdieh for lending me her expertise in eye tracking techniques. Thanks are also owed to Harry Kim, Ivan Collado, Derick Wang, Michelle LeFerve, Shannon Pawloski, Deni Vanpelt, and Jessica De Jong for their excellent work in data analysis. Thanks to Nadine M. and Brandon Pitts, who have been the perfect companions in the lab during evenings (or even late nights) and weekends, for sharing their thoughts in research and life, for cheering me up whenever I needed it, for their generous offers of help, and for making my last semester at Michigan a busy but pleasant time. I'd also like to thank all the labmates who have made this lab the greatest: Sara Lu, Julie Prinet, Katie Tippey, Brian Anthony, Andy Terhune, Richard Ho, Noah Klugman, Khevna Shah, and more.

I would also like to express my deep gratitude to THInC Lab and IOE alumni Tom Ferris, Shameem Hameed, Swapnaa Jayaraman, Dan Nathan-Robert, and Changxu (Sean) Wu for their advice in navigating graduate school, the job market, and more, and for always being there to answer my questions. Thanks are also owed to Shi Cao for being a great classmate and friend, and to other friends in the Center for Ergonomics and IOE for their support and friendship.

I am also deeply appreciative of the participants in my studies for their interest in the research, for always being on time, for their positive attitude in learning such a difficult task, and for giving me valuable feedback. Particularly, I would like to thank those who returned for Experiment 3. Their participation increased the potential broader impact of this research.

I was fortunate to have so many friends along for the journey with me who have always supported me during the long process of completing this project. Specifically, I would like to thank my wonderful housemates at Gamma Alpha: Tara Clancy, Leon Webster, Ya Yang, Katherine Crocker, Chuan Li, and other former and current housemates, for their understanding and emotional support. Thanks also to my former roommates Jie Cheng and Xueying Yu for being my family in my first two years in the US and for making it much easier than it could have been.

My most heart-felt thanks to Samantha Scotland, the most faithful and reliable friend I have ever had. Samantha has been incredibly generous with her time, and I continually draw on her knowledge and experience in both research and life. I appreciate her ‘positive thinking’ in facing any difficult situation and frustration, and her whole-hearted friendship.

Last but not least, I must thank my parents, for their love and support throughout my life. Special gratitude is also extended to my large extended family: my uncles, aunts, and cousins for their understanding, support, and admiration. I am also deeply grateful for the love of Fuyuan Wang, and the support he provided during my pursuit of this degree.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

Introduction

Complex data-rich domains such as space operations, air traffic control, process control and medicine pose significant challenges and impose considerable attentional demands on human operators by requiring them to (1) time share among multiple tasks; (2) monitor a large number of displays; and (3) make decisions as well as select and execute appropriate actions under time pressure and uncertainty. These tasks, at times, exceed operators' cognitive capacity and can lead to breakdowns in performance, thus requiring the introduction of automation technologies that support the operators at various stages of information processing, such as information acquisition, information analysis and integration, action selection, and implementation (Wickens, Lee, Liu & Becker, 2004). The development and introduction of such systems in a wide range of complex domains have been successful in terms of improving performance and reducing the human operator's mental workload during routine operations. However, automation has also led to new, unanticipated challenges, particularly breakdowns in the coordination and collaboration between human and machine and poor joint system performance in case of automation failures and unexpected off-nominal events.

The first problem, breakdowns in human-automation coordination, refers to challenges associated with keeping track of the role and goals of each agent. Many advanced technologies provide poor feedback on automation state and behavior which has been shown to lead to breakdowns in mode awareness (Sarter & Woods, 1995). The second problem, poor joint system performance during automation failures, has been explained by operator complacency or overtrust in the system which results in a failure to detect a problem and intervene or recover when necessary (Crocoll & Coury, 1990; Parasuraman, Molloy & Singh, 1993; Sarter & Schroeder, 2001; Yeh & Wickens, 2001). For example, Galster and Parasuraman (2001) found that experienced pilots detected fewer engine malfunctions when using automated flight deck systems than when performing all flight tasks manually. Also, Metzger and Parasuraman (2005) demonstrated that air traffic controllers were more likely to miss a conflict between two aircraft when the 'conflict probe' automation failed than when they performed the task manually. Both studies reiterate Bainbridge's (1983) assertion that the higher the reliability of the automation, the greater the level of complacency and the greater the likelihood of catastrophic consequences when automation fails. This tradeoff between the benefits of automation during routine tasks and the costs in the case of failures and off-nominal scenarios is critical when determining the proper allocation of tasks and functions to human operators and their automated systems.

The line of research described in this dissertation focuses on analyzing, in detail, the impact of different types and levels of automation on joint system performance. It also examines the effectiveness of different forms of context-sensitive function allocation, namely adaptive, adaptable, and hybrid, as described in more detail in the following sections.

Stages and Levels of Automation

In its early days, automation was conceived of as a dichotomy: it was either on or off. Later, automation was described as varying across a continuum from manual control to fully automated operations. Sheridan and Verplank (1978) first suggested a taxonomy of 10 levels of automation. At the highest level, the automation performs a task autonomously; at an intermediate level, it may choose (and possibly execute) an option unless the human vetoes and intervenes; and at the lowest level, the automation may simply offer recommendations and suggest options to the human operator. More recently, automated systems have been characterized in terms of multiple **stages** (according to the stage of information processing that is supported) and **levels** (ranging from low to high system authority and autonomy) (Parasuraman, Sheridan, & Wickens, 2000; Wickens et al., 1998). In this framework, levels of automation can be defined quasi-independently for four different stages of human information processing: information acquisition, information integration/analysis, decision making, and action execution. A high level of stage 1 automation (filtering information to present only the most relevant) is quite different from a high level at stage 2 (integrating information to diagnose or predict the state of a system), stage 3 (choosing a course of action), and stage 4 (executing the selected action). Higher degrees of automation (DOAs) can be accomplished both by implementing higher levels within a stage and by including later stages. For example, a collision avoidance system that recommends to a pilot to alter the aircraft's trajectory (stage 3) will, by definition, also include stage 2 automation for diagnosing the pending collision in the first place. In turn, in order to accomplish that diagnosis, the system would have collected trajectory data for other neighboring aircraft first (stage 1 automation).

Automation at each stage and level can both benefit and hurt human performance (e.g., Bainbridge, 1983; Wiener & Curry, 1980; Sheridan, 2002). Specifically, high stages and levels of automation reduce workload and increase precision and efficiency; however, because the operator's role is now reduced to monitoring the system rather than actively controlling it (Parasuraman, Sheridan & Wickens, 2008), there is an increased risk of losing awareness of system status and behavior, which can ultimately lead to breakdowns in overall system performance. This tradeoff is particularly prominent when automation is imperfectly reliable and can be expressed as a continuous function within the above stages and levels framework. As the degree of automation increases, workload is progressively reduced and routine performance will continually improve; at the same time, the risk of loss of situation awareness progressively increases. This notion is consistent with empirical findings on automation monitoring and complacency (Molloy & Parasuraman, 1996, Dixon & Wickens, 2006; Bainbridge, 1983) and with very limited empirical data confirming the existence of the above tradeoff (e.g., Endsley & Kiris, 1995; Endsley & Kaber, 1999; Kaber & Endsley, 2004; Kaber, Onal & Endsley, 2000; Crocoll & Coury, 1990; Sarter & Schroeder, 2001).

To date, very few studies have systematically examined the tradeoff space across different stages of automation; fewer still have manipulated both stages and levels or have examined all four dependent variables within a single task. This is rather surprising given that studies of function allocation have been conducted in many different application domains, such as air traffic control (Metzger & Parasuraman, 2005; Sethumadhavan, 2009; Kaber, Perry, Segall, McClernon & Prinzel, 2006), process control (Bahner, Huper & Manzey, 2008; Rottger, Bali & Manzey, 2009), aviation (Sarter & Schroeder, 2001); UAV (unmanned aerial vehicle) control (Calhoun, Draper & Ruff, 2009; Dixon & Wickens, 2006; Cumming & Mitchell, 2007), nuclear power plants (Lin, Yenn & Yang, 2010), military operations (Wright & Kaber, 2005), and spaceflight operations (Lorenz, Di Nocera, Rottger, & Parasuraman, 2001; Manzey, Reichenbach & Onnasch, 2008). Overall, these studies found that a medium degree (i.e., intermediate level and/or stage) of automation is preferable to low or high degrees of automation. But none of the above-mentioned studies systematically varied both levels and stages of automation. A thorough analysis of the effects of joint manipulation of both factors on performance and operator awareness is therefore needed and represents one focus of this dissertation research. Since it is likely that no single degree of automation is optimal for every task and task context, the second critical question addressed in this line of research is the effectiveness of different schemes for adjusting automation stages and levels in a context-sensitive manner.

Adaptive and Adaptable Automation

In the early days of automation, task sets were assigned in a fixed manner to either the human operator or the automation (e.g., Fitt's List, Fitts, 1951). However, operational experience with this approach has shown that it often leads to unbalanced workload and a lack of system awareness on the part of the user (Billings, 1997; Parasuraman & Riley, 1997; Sarter, Woods & Billings, 1997; Sheridan & Parasuraman, 2006). To address these problems associated with inflexible automation (Inagaki, 2003; Parasuraman, 2000; Parasuraman & Miller, 2006; Scerbo, 2001), researchers have identified the need for a more context-sensitive approach to function allocation that varies over time, depending on the human's momentary status and performance, situational factors (e.g., workload), or at the explicit request of the operator (Hancock & Scallen, 1996; Kaber & Riley, 1999; Moray, Inagaki, & Itoh, 2000; Parasuraman, Mouloua, & Molloy, 1996; Rouse, 1988). This context-sensitive dynamic approach can take two forms: (a) 'adaptable automation' where changes in the allocation of functions are initiated by the user, or (b) 'adaptive automation' where the automation triggers changes but the user can override the automation (Barnes et al., 2006; Parasuraman, Bahri, Deaton, Morrison, & Barnes, 1992).

A number of empirical studies on the efficacy of adaptive systems has been conducted over the last decade, in domains such as aviation (Parasuraman, Mouloua, & Hilburn, 1999), air traffic management (Hilburn, Jorna, Byrne, & Parasuraman, 1997; Kaber & Endsley, 2004), industrial process control (Moray, Inagaki, & Itoh, 2000), and human supervision of robots or unmanned vehicles (Parasuraman, Galster, Squire, Furukawa, & Miller, 2005; Wilson & Russell, 2007). Human-in-the-loop simulation studies have shown that, compared to fixed allocations of tasks and functions, adaptive automation can enhance human-system performance (Hilburn, Jorna, Byrne, & Parasuraman, 1997) while reducing potential costs such as reduced situation awareness, complacency, and skill degradation (Kaber & Endsley, 2004; Parasuraman, Mouloua, & Molloy, 1996; for reviews, see Inagaki, 2003; Scerbo, 2007). Parasuraman et al. (2009) found that, in the context of supervisory control of multiple uninhabited vehicles under high workload conditions, change detection accuracy was higher and workload was lower with adaptive automation than with static function allocation and in a manual condition. Similarly, Moray, Inagaki & Itoh (2000) have shown that fault management performance (expressed as the probability of accidents) was better when adaptive automation was employed, compared to a manual condition.

Performance effects of adaptive automation at various stages of information processing were compared in only a small number of studies. Clamann et al. (2002), Kaber et al. (2002), and Kaber et al. (2006) examined the effects of adaptive automation in the context of psychomotor and cognitive tasks and showed that operator performance was better with this type of automation when it was used in the context of lower level functions, such as information acquisition, as compared to adaptive automation in the context of information analysis and decision-making. In particular, Kaber et al. (2006) measured situation awareness (SA) and showed improved SA scores when automation was applied at the information acquisition stage. Clamann & Kaber (2003) compared the performance and workload effects of adaptive automation across four stages of information processing and at two levels of machine authority (suggestions and mandates). Again, performance was superior at the information acquisition stage, as compared to decision-making tasks. And participants were more willing to accept automation suggestions, rather than mandates.

The adaptive (system-controlled) approach to function allocation can employ a number of invocation techniques (Parasuraman et al., 1992) that are based on critical events, operator performance measures, operator's physiological state, models of operator cognition, and combinations of two or more of these criteria. The occurrence of critical events has been used, for example, to change degrees of automation in an air defense system where automation was invoked in response to events such as the activation of a 'pop-up' weapon (Barnes & Grossman, 1985). One disadvantage of basing decisions about the degree of automation on critical events is that it disregards the state of the human operator and may therefore transfer authority to/from the automation when it is not necessary or appropriate to do so. One way to overcome this limitation is to track changes in operator performance instead or in addition to, and use the performance changes to trigger function allocation changes. In this case, operator performance is monitored continuously, and performance degradation is considered a sign of excessive workload, leading to the transfer of tasks to the automation. Mental workload is another possible trigger for re-assigning responsibilities to human and machine. It can be measured using a range of physiological data, an approach that is referred to as 'Augmented Cognition'. To date, various non-invasive physiological (e.g., electrooculograph [EOG], heart rate [HR], galvanic skin response [GSR], blood pressure [BP], and pupillometry) and neurophysiological tools (e.g., portable electroencephalography (EEG), and functional near infarared devices) have been used to assess mental workload in a variety of application domains and environmental conditions (see Reeves & Schmorrow (2007) for a review). The fourth approach to adaptive function allocation uses models of human cognition, which are approximations of human cognitive processes for the purpose of prediction or comprehension of operator state and workload. Rouse & Geddes (1987), for example, based adaptive automation on the human's intentions as predicted from patterns of activity. Preferably, a combination of at least some of the techniques mentioned above should be employed to resolve ambiguities and obtain a more reliable assessment of operator state.

Another approach to context-sensitive assignments of responsibilities to human and machine is adaptable function allocation. The decision whether to use an adaptive or an adaptable approach involves a tradeoff between workload and attentional demands on the one hand and user awareness and perceived control on the other. The adaptable approach to context-sensitive function allocation creates an additional automation management task for the user (Kirlik, 1993). This task will likely have to be performed at times when the user already experiences high workload and attentional demands (Wiener, 1988). For example, Kirlik (1993) developed a UAV simulation in which operators were responsible for manually guiding UAVs to various locations while simultaneously flying and navigating their own (simulated) helicopter. The helicopter could be flown either manually or in autopilot mode, thus representing an “adaptable” task. Results showed that the time required to make decisions about and implement task allocations increased operator workload and hurt overall system performance. This challenge was also illustrated by Bailey et al. (2006) who compared an EEG-based adaptive system and an adaptable system. Their findings show that adaptive automation can enhance situation awareness while reducing mental workload relative to systems requiring user-initiated control. Thus, one challenge for developing an effective adaptable automation system is that the operator needs to be enabled to make decisions regarding the use of automation in a way that does not impose additional workload and interface management tasks.

At the same time, adaptable automation involves a number of benefits. Since the operator is in charge of the assignment of tasks, the result may be improved situation/system awareness, a higher level of (perceived) control over the system, more appropriate automation usage, more balanced mental workload, and increased user acceptance. For example, Wickens (1994) compared the awareness of changes in environmental and system states under self-initiation (adaptable) and agent initiation (adaptive) and showed decreased user awareness in the latter case. In line with the concept of human-centered automation, Miller & Parasuraman (2007) encouraged the use of adaptable automation to ensure that the user remains in control by deciding how much automation to use. They proposed and provided supporting evidence for what they call a “delegation approach” to adaptable human-automation interaction. In this approach, the human operator delegates tasks to the automation at times of their own choosing and receives feedback on system performance. An example of this delegation approach is Playbook™ (Miller & Parasuraman, 2007; Miller, Pelican, & Goldman, 2000). Playbook uses a hierarchical task model to provide a common language with which a human supervisor may communicate goals and intents, and a Hierarchical Task Network planning system

(Erol, Hendler, & Nau, 1994) to understand, reason over, and either critique or complete partial plans provided by the human. It streamlines the process of delegation by providing a compiled set of plans, or “plays,” that are easily invoked and can be further modified as needed (similar to an approved set of plays in sports teams that facilitate task delegation by the team leader). The latter aspect is one critical prerequisite for making sure that adaptable automation does not increase the workload associated with delegation.

A proof-of-concept Playbook prototype has been developed as a mission-planning tool for commanding unmanned combat air vehicles (see Miller, Goldman, Funk, Wu, & Pate, 2004). Also, initial experimental studies of the effects of delegation interfaces on human performance have been carried out (Parasuraman, Galster, Squire, Furukawa, & Miller, 2005). These studies examined the use of a simple delegation interface on system performance during simulated human-robot teaming using the RoboFlag simulation environment. RoboFlag allows the operator the command simulated robots, individually or in groups, at several levels of detail: by providing designated endpoints for robot travel, by commanding higher-level behaviors (or modes or plays) such as “Patrol Border” or “Circle Defense,” or by even higher “super-plays” such as “Go on Offense.” The results from these studies showed that the multi-level tasking provided by the delegation interface allowed effective user supervision of robots, as evidenced by the number of missions successfully completed and the time for mission execution. Further, the increase in mental workload (from workload with manual control) in the flexible condition was relatively small compared to that in the automation control condition.

Potential problems with related forms of context-sensitive automation were highlighted by work comparing a management-by-consent versus a management-by-exception approach to automation control (Olson and Sarter, 2000). In a management-by consent approach, the automation is not allowed to take action unless and until explicit operator consent has been received whereas, under management-by-exception, the automation can initiate actions independent of the operator who has the option to override or reverse system activities after the fact (Billings, 1997). Olson and Sarter (2000) examined pilots' preferences for, and their operational experiences with three different automation management strategies on a simulated modern flight deck: management-by-consent, management-by-exception, and full automation. Pilots expressed a strong preference for management-by-consent; however, in scenarios involving high time pressure, high workload, and low task criticality, pilots preferred management-by-exception. In a subsequent flight simulator study (Olson and Sarter, 2001), the same authors examined the effects of conflict type, time pressure, and display design on pilots' ability to make informed decisions about proposed machine actions in a management-by-consent context. Pilots were asked to fly eight scenarios while responding to a series of air traffic control clearances. Each scenario presented pilots with a conflict that arose from either incompatible goals contained in the clearance or inappropriate implementation of the clearance by automated flight deck systems. Pilots were often unable to detect these conflicts, especially under time pressure, and thus failed to disallow or intervene with proposed machine actions. Performance and verbal protocol data indicate that the observed difficulties can be explained, in part, by poor system feedback.

The above review of benefits and disadvantages and the performance effects of both approaches to context-sensitive function allocation – adaptive and adaptable – suggests that neither approach by itself is sufficient and appropriate. Instead, a hybrid approach may be needed, one which combines the benefits (increased awareness and feeling of control) of these approaches while avoiding their disadvantages (automation management demands). To date, only a few studies have examined such a hybrid approach. The only example is the above described work by Miller et al. (2005) who proposed a form of combined supervisor-subordinate interaction as part of their Playbook architecture. One shortcoming of even this research and other studies on dynamic function allocation is a lack of (information about) interfaces designed specifically to support operators in quickly assessing and adjusting function allocations.

Interfaces in Support of Operator Awareness and Control of System State

One significant challenge for developing effective context-sensitive automation remains the design of associated interfaces that support both at-at-glance awareness of active function allocation settings at varying DOAs and the ability to transition between manual and automated control modes quickly and efficiently. For instance, displays associated with flight management systems on modern flight decks have been shown to provide pilots with inadequate support while they orchestrate and monitor the various degrees of automation at their disposal (Abbott, Slotte, & Stimson, 1996; Sarter & Woods, 1997, 2000). In earlier automation related research, interfaces sometimes failed to include any information about the operator's and the system's responsibilities (Parasuraman, Visser, & Cosenzo, 2009; Caleefato, Montanari, & Tango, 2007); others indicate only whether or not the 'automation is on' without describing its assigned tasks and the degree of automation (Kaber, Perry, Segall, McClernon, & Prinzel, 2006). These interfaces are clearly insufficient for supporting operator's awareness and coordination of function allocation.

Kaber, Riley, Tan & Endsley (2001) summarized some widely agreed-upon requirements for the implementation of interfaces in adaptive automation:

(1) Interfaces should support at-a-glance monitoring and comprehension of system states and function allocations.
(2) Interfaces need to support safe and efficient shifts between different states. When transitioning to manual control, operators need to be provided with an in-depth understanding of how the automation has been controlling the state of the system. When transitioning to automated control, they need information about whether and how well the system has picked up its new responsibility. This information is needed before the actual transition occurs.

(3) The interface needs to be consistent across all possible system states and modes.

(4) Operators need to be informed when they are operating close to their own performance limits or those of the automation.

(5) External attentional guidance should be provided by means of timely and salient indicators in case of acute or pending system changes.

(6) Direct-manipulation interfaces may help avoid or reduce negative effects on primary task performance associated with making changes to system modes and interface features.

These criteria formed the basis for the function allocation interface design that was developed as part of this thesis research and that will be described in Chapter 5.

Application Domain

Teleoperation, i.e., the communication with, and control of, a remotely located machine by a human operator, is widely used in domains that require the performance of tasks in environments that are dangerous or inaccessible to human operators. Examples of such domains include medicine, underwater exploration and space operations. In the case of space exploration – the application domain for this research - diverse tasks such as the deployment of satellites, service for the Hubble Space Telescope, maintenance of payload, inspection and repair of the Space Shuttle or the construction of the International Space Station have been performed through teleoperation, using either the Space Shuttle or the ISS Remote Manipulator Systems (RMS, see Figure 1.1) (Brandan, M.A.M., 2007).

Control of robotic systems in microgravity can extend over long periods of time (in some cases, more than 7 hours) and represent a delicate and risky activity. The complexities of these tasks are related to the need to handle large masses (> 100,000 kg), multi (6- and 7-) degree-of-freedom systems, unique end effectors (EE), limited possibility of direct visual inspection, a wide range of manual and automatic control conditions, and high risk. Safe and efficient robotic operations require that the crewmembers possess unique coordination and manipulation skills and detailed knowledge of the system's design and operation (Currie & Peacock, 2004).

Learning to manipulate these systems requires many hours of training and practice. On the International Space Station, for example, there are no windows near the Robotic Workstation (RWS) in the Destiny Laboratory module; here, arm operators depend on three monitors to provide camera views of their workspace. Manipulating the arm is relatively intuitive if the cameras are oriented in the same way as the arm's control frame (i.e., if the operator moves the arm to the left, the arm moves to the left in the camera view), but when the orientations of the cameras and the control frame do not coincide, the required mental transformations make controlling the arm a highly complex and demanding task.

Operators of the robotic arm on the Shuttle or Space Station must constantly be aware of the spatial location, configuration and motion of all components in their workspace. This includes not only segments and joints of the arm, but also payloads, structures, and astronauts during extra vehicular activites (EVAs). Maintaining this awareness, mentally integrating the information provided by each camera view, and adjusting the camera views in real time are demanding tasks. Therefore, operators most often work in pairs. The primary operator is responsible for actually manipulating the arm using the hand controllers while the secondary operator observes the task and assists with camera or mode adjustments, clearance and singularity monitoring, and task planning (Tomlinson, Z. A., 2009).

The current robotic arm on the International Space Station employs a rather limited set of automation modes: 1) the 'frame of resolution' mode, in which the operator enters the target location and the system then moves the arm automatically to that target, and 2) the 'joint automatic' mode, in which the operator enters the target joint angles of the arm, and the system then operates the arm to achieve the desired joint configurations. Even during automatic operations, the operators still have to conduct the same planning activities as in manual mode. The planning tasks include selecting the proper camera view, lighting, and frames of reference. The operator also has to continuously monitor the system while it is operating the arm. The current system is not able to recognize obstacles (e.g. a structure, a payload) and thus does not provide any collision avoidance/warnings.

Figure 1.1: Robotic arm (left) is controlled by the astronauts at the robotic workstation (right) inside International Space Station (source: RLROUSE.com, nasa.gov)

In the following chapters, different types of space teleoperation automated aids are designed and evaluated through simulation studies to examine the performance effects of levels and stages of automation, and the effects of dynamic function allocation schemes. In Chapter 2, we developed visual and tactile warnings to better support operators in monitoring the arm configuration and position, and consequently avoiding collisions and singularities. The findings from Chapter 2 informed the design of automation aids described in Chapter 3, where a study investigated the effects of different levels of automation for supporting all four stages of information processing represented by hazard avoidance, camera selection, and trajectory control. In Chapter 4, we describe a meta-analysis of 14 studies that provided integrated insights on the topic of effects of levels and stages of automation. Both the findings from the above study and meta-analysis, as well as operational experience with current automation technologies, highlight the need for context-sensitive use of automation. In the study described in Chapter 5, three different dynamic function allocation schemes, i.e. adaptive, adaptable, and hybrid schemes, were comparatively evaluated.

Chapter 2

The Design and Evaluation of Visual and Tactile Warnings in Support of Monitoring the Arm Configuration (Stage 1 Automation)

Introduction

One of the main human factors challenges associated with the use of RMSs is the need to support operators in maintaining awareness of the arm position and configuration. This is critical to avoid problems such as singularities and joint limits. A singularity is a configuration in which the arm loses one or more degrees of freedom. The three singularities that the simulated robotic arm in this experiment can encounter are: (1) elbow pitch singularity (elbow pitch angle is 0 degrees or +/- 180 degrees), (2) wrist yaw singularity (wrist yaw angle is +/-90 degrees), and (3) wrist over shoulder singularity (the wrist pitch joint is directly above the shoulder of the arm; see Figure 2.1 for the examples of these three singularities). Joint limits are defined as the joint angle at which the end of the mechanical range of the arm is reached. Both joint limits and singularities can cause the arm to stop abruptly and result in hazards and unsafe conditions.

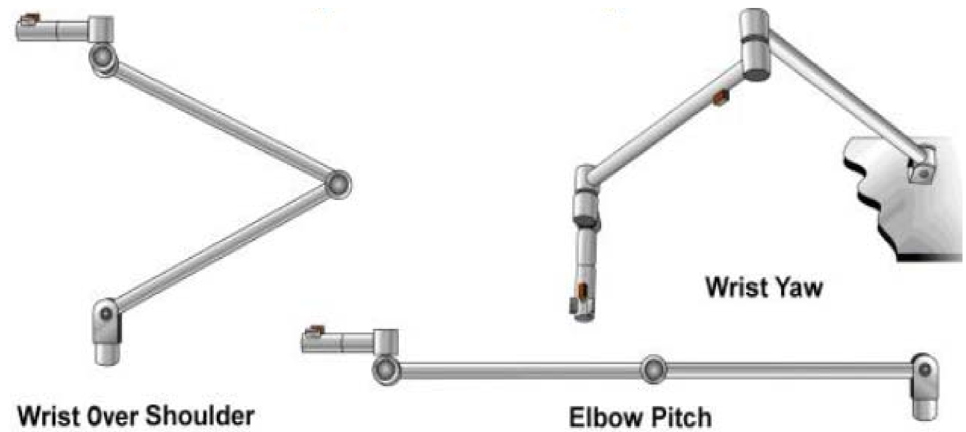


Figure 2.1: Arm singularities in BORIS (source: GRT Handbook, NASA Johnson Space Center)

Several factors contribute to difficulties with maintaining awareness of the arm configuration. First, the current robotic arm interface shows only the current angle but does not provide reference points, such as the problematic ranges for singularities and joint limits (see Figure 2.2 for the joint angle indicators on the current Arm Control Graphic user interface (GUI)). Thus, it violates the principle of putting 'knowledge in the world' (Norman, 1988) and imposes undue memory demands on the operator. Also, the display is not designed to support at-a-glance checking since the problematic joint angles for elbow and wrist singularities are not aligned consistently. The elbow pitch singularity (at 0 degrees) is located at the center of the bar, whereas the wrist yaw singularities (at +/ – 90 degrees) are reached close to the two ends of the column.

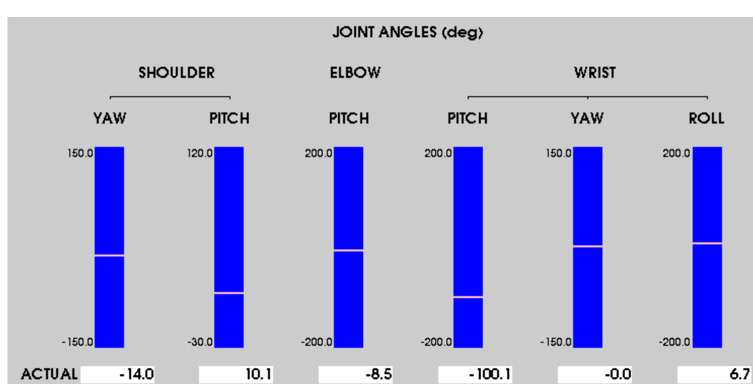


Figure 2.2: Original joint angle indicators

Even with an improved design of the joint angle display, another problem with monitoring the arm configuration is the need for astronauts to divide their attention between multiple visual tasks and numerous displays/task views that are distributed across two monitors. During training, astronauts and participants in our studies are taught a hub-and-spoke scan pattern in which the operator's attention regularly moves back and forth between a central task view (hub) and other surrounding views and the GUIs. However, they still tend to focus largely on the task view on the right-hand monitor, especially when they need a close-up view of the arm and/or end effector. For example, Figure 2.3 shows eye tracking data for one participant in our research for a regular 'fly-to' task that illustrate this problem. Note that very few fixations were in the area of the joint angle display on the left-hand monitor.

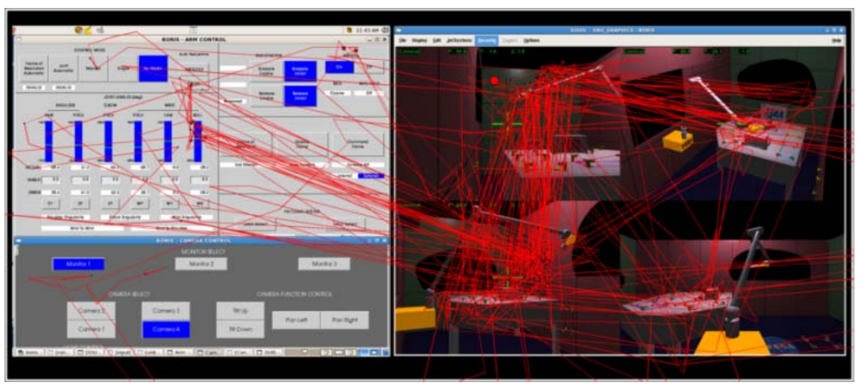


Figure 2.3: Typical eye tracking data (fixations) for one participant during a 'fly-to' task

A possible solution to this problem is to reduce scanning costs (Wickens, Gordon, & Liu, 1998) by integrating important arm configuration information with the view of the arm and workspace on the right-hand monitor. For example, when the arm is approaching a problematic configuration, color coding can be used on the corresponding segment of the arm to warn the operator of a pending singularity or joint limit. This approach complies with the proximity compatibility principle (Wickens & Carswell, 1995). The arm configuration and the location and movement of the arm must be mentally integrated by the operator to complete the task, i.e. they have "mental proximity". The interface should support this integration through “display proximity” by displaying both pieces of information close to each other or even superimposed.

Finally, given the significant visual attentional demands imposed by the arm control task, a third promising means of supporting astronauts in monitoring the arm configuration could be to present warnings of pending singularities and joint angles via non-visual signals. Tactile information presentation is particularly effective for providing spatial guidance and notifications and has been proven successful in other domains, such as aviation (e.g. Sklar & Sarter, 1999), driving (e.g. Ho, Tan, & Spence, 2005) and medicine (e.g. Ferris & Sarter, 2009). Tactile cues have also been used to support interaction with objects (MacLean & Hayward, 2008), help with orienting and guiding in 2-D or 3-D space (e.g. van Erp, 2001, 2005), direct attention to a visual target on a screen (Tan, Gray, Young &Taylor, 2003), and navigate unfamiliar terrain (Jones et al., 2006). The ability to localize vibrotactile stimulation on the body is best supported by presenting it near anatomical points of reference such as the wrist, elbow, spine, or navel (Cholewiak & Collins, 2000; Cholewiak, Brill, & Schwab, 2004; van Erp, 2001).

The goal of the present study was to comparatively evaluate the effectiveness of the above three approaches: (1) the improved GUI on the left-hand monitor, (2) integration of configuration information with the arm view using visual highlight, and (3) tactile notifications for better supporting monitoring of the arm configuration and for avoiding joint limits and singularities. The best performance was expected for the tactile warnings, followed by visual highlight and integration and finally the improved GUI only.

Method

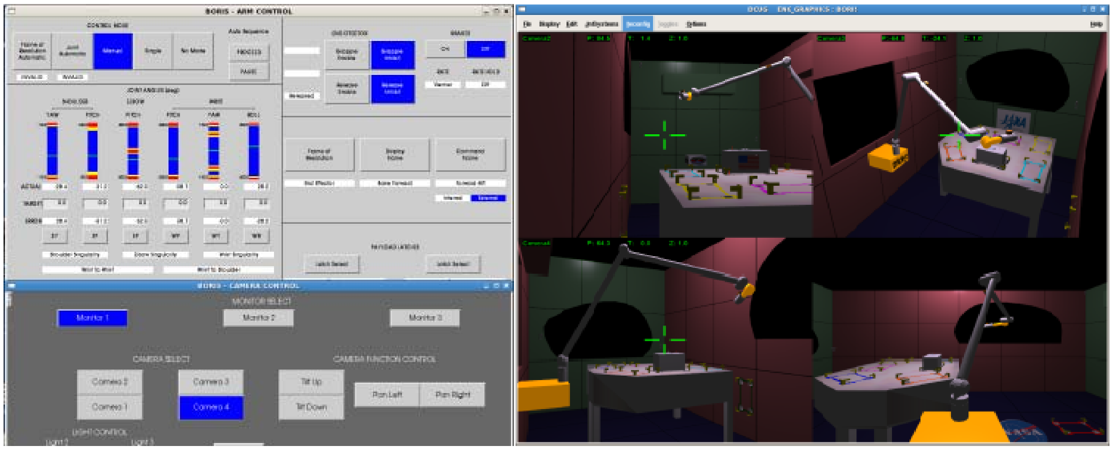
Participants

Twelve University of Michigan graduate and undergraduate students participated in the experiment (7 males and 5 females; mean age = 23.9, SD = 3.5). They reported normal or correct-to-normal vision and no compromised sense of touch. None of them had previous experience with operating a robotic arm. The participants were paid $15/hour for their participation.

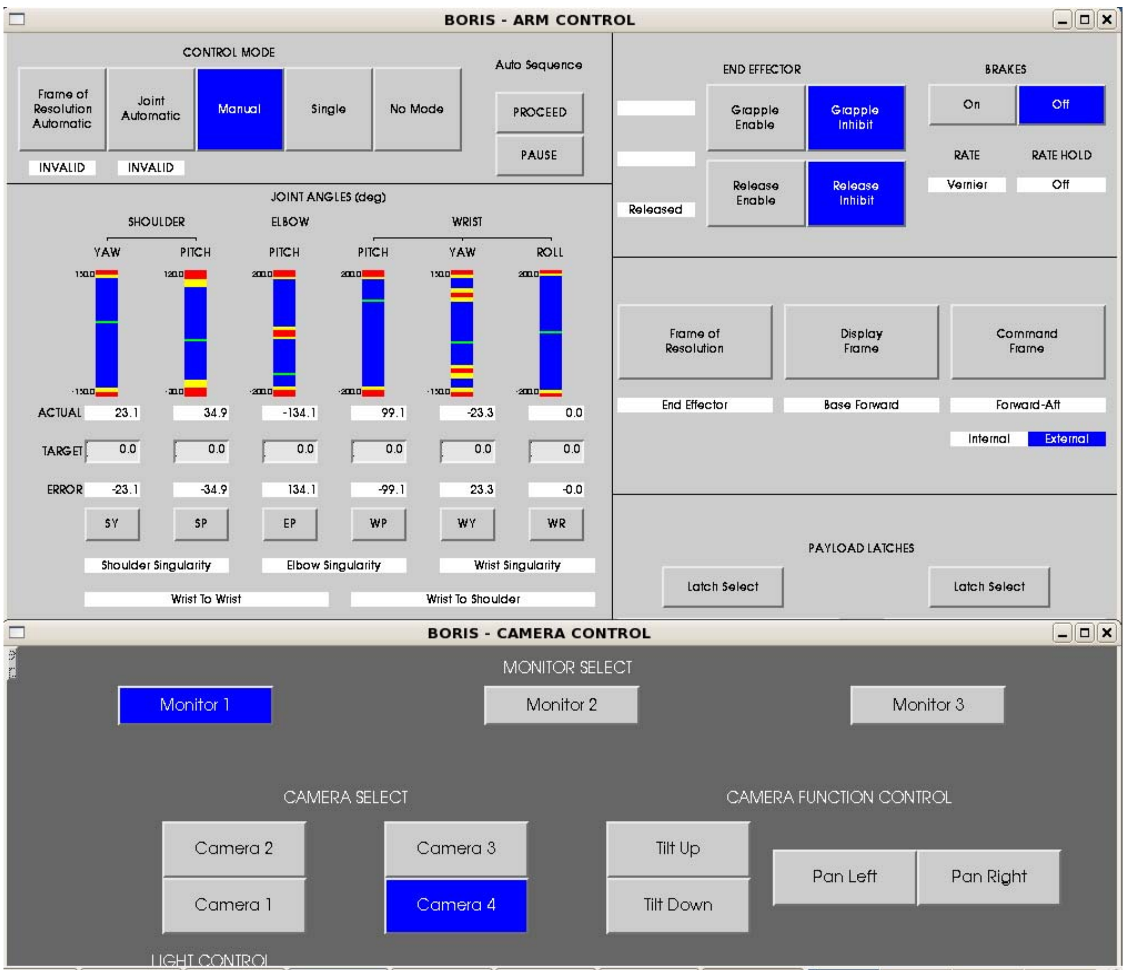
Robotic Arm Simulator

NASA's simulated robotics environment, BORIS (Basic Operational Robotic Instructional System), was duplicated as a desktop simulation in our laboratory. The BORIS virtual environment consists of a rectangular room that contains the robotic arm, which has six degrees of freedom: shoulder pitch, shoulder yaw, elbow pitch, wrist pitch, wrist yaw, and wrist roll. It also includes a table for practicing payload manipulations (see Figure 2.4). A number of different task views are available. Cameras are located at each corner of the room, at the arm's elbow joint, and on the end effector (a device at the end of the arm that can be used to manipulate payload) of the arm. The cameras on the walls have adjustable pan, tilt and zoom control, while the cameras on the arm have zoom control. In addition, there is a view from a window located behind the base of the robot arm. While operating the arm, operators can display simultaneously as many as three camera views, plus the window view, on one of two 19-inch monitors.

On the other, left-hand monitor, two panels are shown, including 1) the 'Arm Control GUI' for selecting command modes and coordinate frames, and, more important for this study, the arm configurations; and 2) the 'Camera Control GUI' which supports the selection of cameras, lights, crosshairs and rate overlays (see Figure 2.4 below).



(a)



(b)

Figure 2.4: (a) BORIS virtual environment. Upper left: Arm Control GUI. Bottom left: Camera Control GUI. Right: camera and window views. (b) Zoom-in view of GUIs on the left-hand monitor

Tasks

Each participant played the role of a payload specialist in a simulated space mission and was responsible for operating the simulated robotic arm. Participants were asked to perform fly-to tasks, i.e. they had to move the robotic arm from a given starting point to a target location and orientation. This task involves five steps:

(1) Selecting 3 camera views, which together should serve three purposes including (a) the big picture view, (b) the task view (to monitor if the arm moves as expected or to the target location and orientation, this step requires drawing on a floor plan of the room), and (c) the clearance view (to monitor whether there is a potential collision between the arm and a structure.

(2) Selecting frames of reference to define the location and the orientation of the arm. Three types of frames are used to define the location and the orientation of the arm, and the relationship between hand controller inputs and the movement of the arm: a display frame fixed to a structure, a frame of resolution (FOR) fixed to an object, e.g. a payload or the end effector of the arm, and a command frame that defines the relationship between hand controller inputs and the movement of the arm. The command frame has two options: external or internal command frame, aligned with a fixed structure or an object respectively.

(3) Selecting the rate of arm movement (Coarse - a faster rate - vs. Vernier - a lower rate) according to the flight rule (switch to the vernier rate when the arm is within 1.5m of structures).

(4) Selecting the control mode (in this experiment, ‘manual mode’ was used for all sessions).

(5) Operating the arm with two hand controllers. The translational hand controller (THC) controls the movement along X, Y, and Z axis; the rotational hand controller (RHC) controls rotations about 3 axes (pitch about Y axis, yaw about Z axis, and roll about X axis). In this experiment, the arm was controlled by two 3-axis joysticks. The RHC was a Logitech Extreme3DPro USB game controller with 3 axes

(right/left, up/down, twist). The THC was modified from a RHC and has right/left, up/down, and in/out.

While operating the arm, the operator has multiple sub-tasks, including:

(1) Recall the plan for hand controller movement and actually move it along multiple axes at once.
(2) Recall the target location and the alignment of the end effector with a fixed structure at the target location.
(3) Monitor the camera and window views to: a) make sure the arm moves as expected; b) gain a sense of the arm's location and orientation; c) monitor the clearance between the arm and structure; d) monitor for potential singularities, joint limits, and self-collisions.
(4) Monitor the arm control GUI on the left-hand monitor to gain a sense of the joint angle and the distance to joints limits. The angle range and current angle for each joint is shown on a blue bar. Located below the joint angle displays are the displays for singularity and self-collision. For each case of singularity, for example, there is an indicator which turns red when the arm has already reached a joint limit (see Figure 2.5(a)).

Three fly-to tasks were designed that involved a high likelihood of reaching either a singularity or a joint limit.

Experimental Design

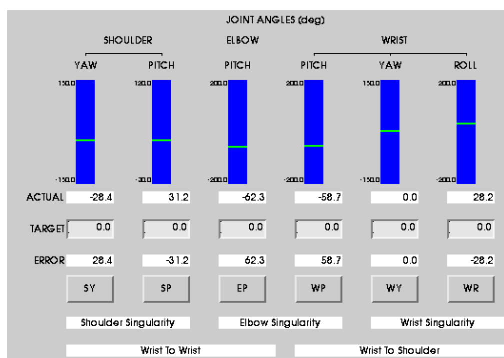
The independent variable in this study was the type of support provided for monitoring the arm configuration. The first condition was an improved 'Arm Control' GUI (see Figure 2.5(b)) where color-coded reference markers and warning ranges were included. The angle range in which the arm would be stopped by the system ('stopping range') was marked in red. The angle range 10 degrees outside the stopping range was marked in yellow (this range is called 'warning range' for the remainder of the paper). The stopping ranges and warning ranges for all singularities and joint limits are shown in Table 2.1. This change to the original GUI was intended to support at-a-glance monitoring of the arm configuration and quick detection of impending problems on a dedicated display. This condition is the baseline condition in this experiment.

Table 2.1: Warning and stopping range for each joint angle (degree)

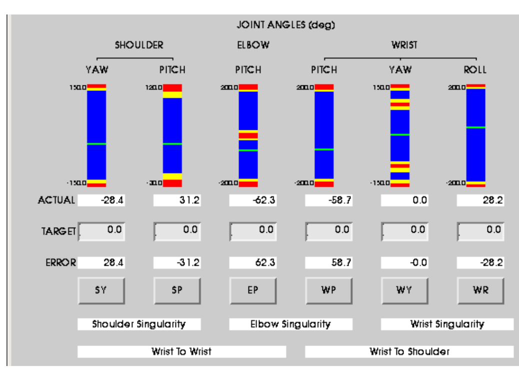
Problem		Yellow range Arm turn yellow Tactor low intensity	Red range (stop) Arm turn red Tactor high intensity
Singularity	Wrist yaw	[80, 85], [95, 100] [-85, -80], [-100, -95]	[85, 95] [-85, -95]
	Elbow pitch	[-20, -10], [10, 20]	[-10, +10]
	Wrist over Shoulder	---	not defined by joint angle
Joint Limit	Shoulder Yaw	[130, 140] [-140, 130]	[140, 150] [-150, -140]
	Shoulder pitch	[100, 110] [-20, -10]	[110, 120] [-20, -30]
	Elbow pitch	[170, 180], [-180, -170]	[180, 200], [-200, -180]
	Wrist yaw	[130, 140], [-140, -130]	[140, 150], [-150, -140]
	Wrist roll	[180, 190], [-190, -180]	[190, 200], [-190, -200]
	Wrist pitch	[170, 180], [-170, -180]	[180, 200], [-180, -200]

In the second condition, the information regarding problematic arm configurations was not only color-coded on the GUI, but was also integrated with the view of the arm itself in the various camera views. When a problematic configuration was approached, i.e. the arm was entering a warning zone, both the corresponding joint segment and the base of the arm turned yellow (see Figure 2.5(b)). When the arm entered the stopping range, the joint segment and the base turned red. This condition ('visual highlight') was expected to attract attention to problematic configurations and minimize scanning costs.

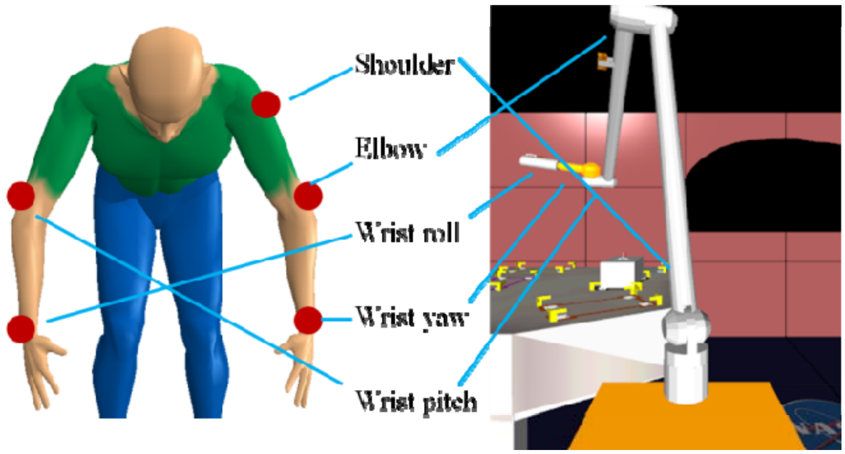
The third condition was a tactile display of the warning zones and stopping zones. Vibro-tactile signals were presented via five tactors (C-2 tactors developed by Engineering Acoustics, Inc., oscillate a 7.5-mm "skin contactor" at frequencies near 250Hz with a displacement of approximately 1mm). These tactors were arranged on the participant's arms, secured via two elastic sleeves (for wrist and elbow) and one elastic shoulder belt (see Figure 2.5(c)). The arm was chosen as the presentation site to create a natural mapping between robotic and human arm joints. Three tactors were used to provide warnings related to the robotic arm's three wrist joints - wrist pitch, wrist yaw and wrist roll. Because little is known about humans' ability to distinguish reliably between tactile stimuli applied to multiple locations on the wrist (Oakley, et al. 2006, Chen, et al, 2008) and because frequent wrist flexion and extension while controlling the hand controllers makes it difficult to maintain a fixed mapping between tactor location and wrist joint, the tactors were mapped as follows in the present experiment: wrist pitch - elbow on the right arm, wrist yaw – left wrist, and wrist roll - right wrist. The tactors vibrated at a frequency of 180Hz when the corresponding joint angle entered the warning range and at a frequency of 250Hz when the joint angle approached the stopping range. The tactile warnings were intended to off-load the visual channel while providing spatial information regarding problematic arm configurations via natural mappings. In this condition, the improved GUI on the left-hand monitor was presented but not the visual highlight on the robotic arm itself.



(a) Original joint angle indicators



(b) Improved 'Arm Control' GUI



(c) Mapping of joints in visual highlight and tactile

Figure 2.5: Displays to support operators in maintaining awareness of the arm configuration: (a) Original joint angle indicators; (b) Improved 'Arm Control' GUI; (c) Mapping of joints in visual highlight and tactile warnings

The dependent measures in this study were reaction times to warnings (or entering a warning zone), the frequency of encountering problematic arm configurations, scan patterns (eye fixations in the area of joint angle displays), and subjective preferences for the different warning presentation conditions. Eye tracking data were collected using the ASL Eye-Trac 6D eye tracking system (sampling rate 60 Hz).

Procedure

Participants first attended two 2.5-hour training sessions on two consecutive days to learn about the concepts and skills required for operating the arm. During the training, the participants performed 3-4 fly-to tasks with the first warning presentation condition. All participants were able to complete the training successfully, i.e., they completed the last fly-to task within tolerance limits (2 meters for translational movement along any axis, 30 degrees for rotations about any axis). On the third day, they completed three experimental runs. Participants performed one fly-to task under each of the three warning conditions. To counteract learning effects, the order of the experimental conditions, and the tasks associated with each condition, were counterbalanced between participants. After the experimental sessions, the participants completed a debriefing questionnaire to indicate their preferred warning condition, the strategies they used to perform different steps of the tasks, and any other comments and suggestions about the designs and experiment.

Results

The data were analyzed with repeated measures ANOVAs using general linear models (formulated in IBM SPSS 19). All participants completed the scenarios within tolerance. Therefore, the analysis of performance measures focused on reaction time and the frequency of problematic arm configurations. Performance and eye movement data were compared for the three conditions.

Reaction Time

Participants responded to warnings by either releasing the hand controller(s) or by reversing the hand control input along one or more axes. Therefore, reaction time was defined as the time between when the arm entered a warning area and when the hand controller inputs became zero or changed to the opposite sign (from + to -, for example). There was a significant difference between reaction times in the three conditions ($F(2, 22) = 4.175$, $p = 0.029$; see Figure 2.6). Post-hoc comparisons showed that reaction time in the tactile condition (0.59s) was significantly shorter than in the improved-GUI-only condition (1.39s; $p=0.040$) and the highlight condition (1.06s; $p = 0.041$).

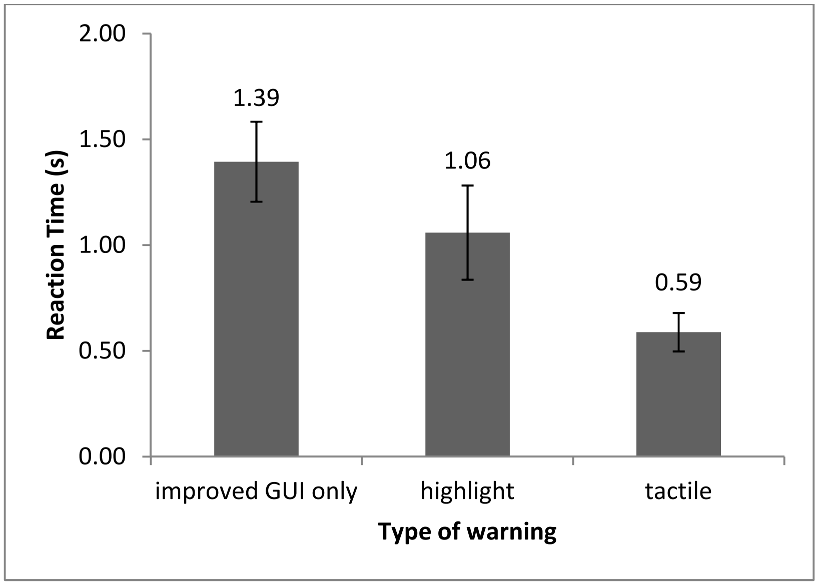


Figure 2.6: Average reaction times to warnings as a function of warning type (error bars represent standard errors)

Frequency of singularities and joint limits

In some cases, participants failed to notice a warning in time and, as a result, they failed to stop the arm from moving into a singularity or a joint limit, i.e. the joint angle entered a stopping zone. The average number of times the arm entered a stopping zone differed significantly as a function of warning condition (F (2, 22) = 5.098, p = 0.015; see Figure 2.7). Post-hoc tests showed that they were significantly more likely in the improved-GUI-only condition than in the tactile condition (p=0.011), and marginally more likely in the highlight condition (p=0.059).

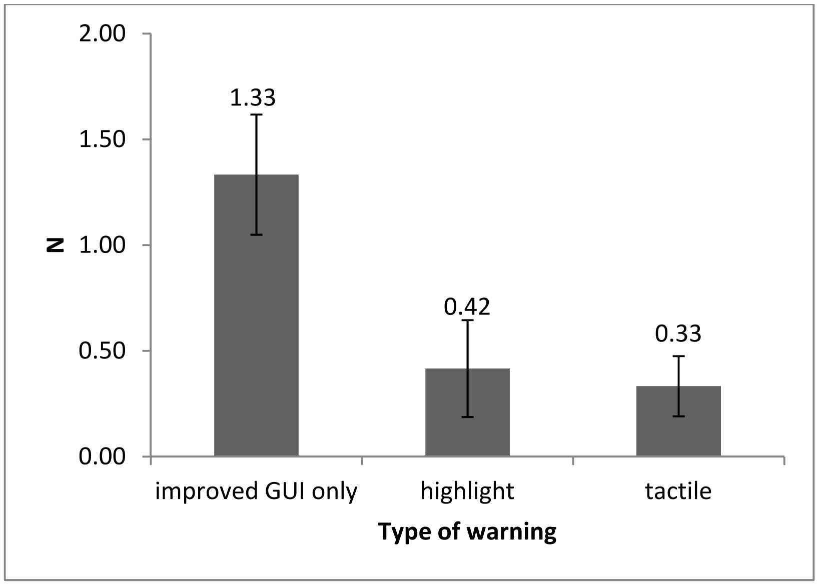


Figure 2.7: Frequency of entering stopping zone as a function of warning condition. N= number of times the arm reaches a singularity or joint limit in one scenario. Error bars represent standard error

Scan Pattern

Eye tracking data from five participants had to be excluded from the analysis due to either low quality of the signal (in one case because the participant was wearing glasses, in another case because of large head movements) or because of calibration issues. The scan patterns of the remaining 7 participants were analyzed in terms of the percentage of dwells (defined as fixations of a duration of more than 1 second within an area of interest) on the arm configuration indicators on the Arm Control GUI.

There was no significant effect of warning condition due to large inter-individual differences. However, trends were observed suggesting more fixations on the Joint Angles Display in the improved-GUI-only condition (mean = 4.49%, SD = 3.19%) than in the tactile condition (mean= 3.69%, SD = 1.58%). The smallest number of fixations was observed in the visual highlight condition (mean = 2.83%, SD = 1.60%).

Subjective Rating

In the debriefing questionnaire, participants were asked to rank the three conditions in terms of efficacy of supporting arm monitoring. Six participants (50%) preferred the tactile warning condition whereas the other six subjects (50%) ranked the visual highlight condition as being most effective. They were also asked about their overall preferences for the three conditions. Seven people preferred visual highlight, three preferred tactile, and two did not respond.

According to participants' comments, the tactile warnings were preferred because they reduced visual attentional demands (e.g., "I was able to focus all my attention to the cameras") and were effective in preventing critical events (e.g., "…was relatively free from the risk of running into any singularity"). Reported problems with tactile warnings include intrusiveness (e.g., "vibration was too intrusive", "tactors startle me and can be very distracting/confusing"), unfamiliarity (e.g., "I was not used to the tactors and felt a little bit nervous when it was suddenly on", "Hard to tell where the problem was"), and short duration (e.g., "the highlighting remained on screen while the tactors would simply stop vibrating and one would simply have to remember where it occurred").

Favorable aspects of the visual highlight included an intermediate level of salience/less disruption ("It did not surprise me like tactors, but I was able to catch the singularities" or "It allows me to continue working but I still know a constraint was being reached") and intuitiveness.

Discussion

Maintaining awareness of the robotic arm's configuration during space teleoperations continues to be a major challenge for astronauts. This study aimed at comparing the effectiveness of three different approaches to supporting operators' awareness of the arm configuration and pending problems, such as joint limits and singularities.

Both visual highlight (information integration) and tactile warnings led to improved operator performance. Participants were less likely to approach the stopping range than with the improved GUI alone. This could be explained, in part, by the fact that they participants responded more quickly to entering a warning zone (getting close to singularity/joint limit, in the GUI only condition) by eliminating the hand control inputs. It should be acknowledged that the observed performance benefits may be attributable to the mere presence of a warning rather than the particular types of warning employed in this experiment. Reaction times with the tactile warnings were shorter even than in the visual highlight condition. However, one problem with the tactile display is the discontinuity of the signal. Due to the limitation of the tactors, the tactile warnings were presented as three pulses upon entering a warning zone or stopping zone. Participants stated that they sometimes had to recall which one the warning was before making a judgement. They would prefer the warning being present at all time while the arm is in a warning zone.

The tactile display was also confusing at times because warnings were presented to both arms and the participant found it hard to locate the problem. One solution could be to combine the visual and tactile warnings. Thus the advantages of both displays (visual is more intuitive, tactile is faster) could be combined – a 'best of both worlds' approach. In the combined approach, the arm joint will be highlighted on the camera views, while the tactile warning will also be concurrently provided upon entering a warning zone.

Interestingly, the two warning conditions had a positive impact on reaction times and problem avoidance but may have also resulted in an unexpected and undesirable operator behavior. It was expected that operators would scan the Arm Control GUI in response to warnings to increase their awareness of the arm configuration. Instead, the eye tracking data show that the participants scanned the Arm Control GUI less often when warnings were provided. This suggests that, instead of trying to use available information to anticipate and avoid singularities and joint limits on their own, operators may have come to rely on the warnings instead, which is a good example of automation complacency. Note, however, that the observed overreliance on warnings could also be attributable to participants' (i.e., students') lack of experience with teleoperation tasks.

It was also expected that participants would reduce the intensity and/or change the direction of hand controller inputs accordingly to avoid entering the stopping range. Instead, participants tended to release the hand controllers. While this avoids entering the stopping range, undesirable 'pauses' and abrupt arm movements are introduced. This problem can likely be addressed through more extensive training and explicit instructions for how to handle warnings.

The eye tracking data showed that the participants still needed to refer back to the GUI on the left monitor to gain more specific information about the joint angles. This implies that more work is needed to develop automated aids that support later stages of information processing, beyond notification of a problem (Parasuraman, Sheridan, & Wickens, 2000). For example, when a warning is presented, a more detailed diagnosis of the problem and suggestions for appropriate hand control inputs could be displayed also.

The findings from this experiment form the foundation for the design of hazard avoidance automation in Chapters 3 and 4.

Chapter 3

Effects of Stages and Levels of Automation on Performance, Workload, and Scan Pattern

Introduction

As mentioned earlier, automation can both benefit and hurt performance of a joint human-machine system (Bainbridge, 1983; Sarter, Woods, & Billing, 1997). It has been shown to lead to issues such as a loss of mode awareness (e.g. Sarter & Woods, 1994), automation complacency and bias (e.g. Metzger & Parasuraman, 2005; Parasuraman, Molloy, & Singh, 1993; also see Parasuraman & Manzey, 2010, for a review), and vigilance decrements.

To overcome these problems, it is critical to ensure that the appropriate level of automation (LOA) is used for a given automation function at the related stage of information processing it supports (Parasuraman, Sheridan, & Wickens, 2000). For each stage (information acquisition, information analysis/integration, decision making/action selection and action execution), automation can act at various levels of autonomy and authority.

Previous research has studied the effects of LOA on joint system performance and operator acceptance but no consensus has been reached. For example, Endsley and Kiris (1995) investigated LOA effects in the context of a decision-support system for a navigation task. They found advantages of medium levels of automation for preventing effects of skill loss, reflected in poor manual performance on a decision task after prolonged use of automation support. Intermediate levels of automation have also been shown to improve situation awareness and reduce workload in supervisory control tasks (e.g. Ruff et al. 2002). However, another study investigating LOA effects on fault identification and management in a simulated process control (Lorenz, Di Nocera, Roettger, & Parasuraman, 2002) found that higher levels of automation were associated with better performance when operators had to abandon the automation and return to manual task performance.

To date, few studies have examined LOA effects in the context of space operations (Di Nocera, F., Lorenz, B., & Parasuraman, R., 2005). Furthermore, to our knowledge, there are no studies that have compared LOA effects for different stages or types of automation in a single study. To fill this gap, the current study examined LOA effects in both routine and automation failure scenarios for three stages of automation during a space tele-operation task.

A cognitive task analysis revealed that three aspects of this task would potentially benefit from the availability of an automated aid. The first sub-task is the monitoring of the arm configuration and the clearance between the arm and surrounding structures. As discussed in Chapter 2, operators tend to focus on monitoring the arm movement and rarely scan the 'Arm Angle Display' to gain information about the arm configuration. Hazard avoidance was thus further developed based on experiences with the visual warning in Chapter 2. This function was enhanced to include both collision and self-collision avoidance because clearance between the arm and surrounding structures was shown to be difficult for operators to judge while moving the arm. Hazard avoidance automation is considered stage 1 (obtain joint angle information) and stage 2 (process arm position information to judge clearance) automation.

The second subtask is camera selection, which includes the initial selection of a camera view before moving the arm and also switches to different views during arm operation. The latter was particularly difficult for participants because, while operating the arm, the operator tends to focus on watching the arm movement from one of the camera views and thus fails to pay attention to other views, a failure of divided attention. As the arm continues to move across the room, an initially good camera view might become obsolete. For example, the end effector (EE) may no longer be visible; the EE movement may become difficult to monitor if the camera angle is now parallel with the EE movement; or there may no longer be any clearance view at all. The participant may not adjust the camera view until a critical event occurs, such as the arm approaching a structure. Thus camera recommendations were developed to assist operator in selecting and changing camera views. The recommendations were based on a computational model called FORT (Frame of Reference Transformation; Gacy et al., 2011). In FORT, each camera is assigned a penalty based on (1) the distance between the camera and the EE (the further apart, the higher the penalty), (2) the relative angle between the current direction of arm movement and the orientation of the camera (the penalty was smallest if the two were perpendicular), and (3) the visibility of the EE in the camera.

The third subtask that was expected to benefit from some form of automated aid is actually controlling the arm to fly a planned trajectory with optimal hand control inputs. Operators tend to spend a long time trying to determine the appropriate hand control input, particularly with respect to the rotational hand controller. They also had to remember the optimal path and try to determine the best multi-axis hand control (e.g. both x and y direction) for this envisioned trajectory. This added to their already high workload, and it was expected that their performance would benefit from a visible reference that was integrated with the camera views.

Overall, it was expected that increasing levels of automation would result in progressively better performance during routine operations but incur performance costs during automation failures. Specifically, higher levels of hazard avoidance should lead to increased safety as the frequency of encountering (near) collisions would be decreased. Second, trajectory control automation was expected to shorten trial completion time by helping the operator fly a more efficient path (using multi-axis hand control as much as possible, with the correct ratio of inputs along two axes) and thus reduce deviations and workload. Third, the availability of camera recommendations was expected to improve the quality of camera selections and thus improve monitoring and EE control performance, compared to manual camera selection.

Method

Participants

Thirty-six University of Michigan undergraduate and graduate students participated in this experiment (23 males and 13 females; mean age = 22.9 yrs, SD=3.3). They reported normal or corrected-to-normal vision (only contact lenses were acceptable due to the use of an eye tracker). None of the participants had any previous experience with operating a robotic arm. Participation was voluntary, and compensation of $15/hr was paid for completion of the training and experiment sessions.

Robotic Arm Simulator

The simulator used in this experiment was again BORIS (same as in Chapter 2). Two changes to the interface were made: (1) The right-hand monitor presented slightly different views from those described in Chapter 2. In particular, in the left bottom corner, instead of presenting a camera view, a digital map showed a top-down view of the room and the target location. Lines connecting the cameras and the target were presented on the map. The digital map and the lines were supposed to match the participant's drawing on the paper-based map for planning purposes. It was presented to minimize the need to look at the paper map, thus reducing scanning costs for the participant and enhancing accuracy of the eye tracking data. (2) Another difference was a new GUI, the 'Automation Target' GUI, which was added and placed next to the ‘Arm Configuration GUI’. Participants used this new GUI to enter the coordinates of the target point when they were flying the arm with trajectory guidance or using auto trajectory control. Figure 3.1 shows the layout of the displays and controls across the two monitors.

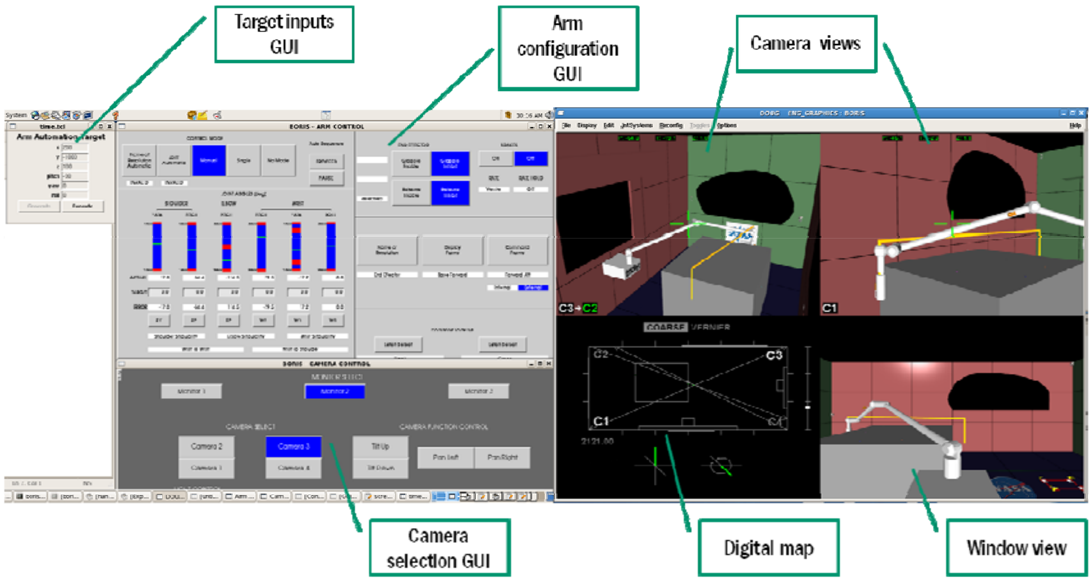


Figure 3.1: Layout of camera/window views and visual displays in BORIS virtual environment

Tasks

Each participant played the role of a payload specialist in a simulated space mission and was responsible for operating the simulated robotic arm. Participants were asked to perform fly-to tasks with the goal of moving the end effector (EE) to a given target location and orientation within tolerance (1 meter for positions, 15 degrees for orientation). The steps involved in performing the task were similar to those described in Chapter 2, with two exceptions. First, the participant chose two camera views from the four cameras in each corner of the room (instead of choosing three from six). The two cameras on the arm (one on the EE, the other on the boom) were not listed as options because: a) the cameras located in the four corners of the room were considered sufficient for fly-to tasks, and b) the camera recommendation algorithm only calculated the penalty scores for these four cameras. The third window that originally presented a camera view in BORIS was used for a digital map.

Second, the participants did not need to choose the frames of reference which were preset as follows: the display frame was set to 'Base-FORWARD', the frame of resolution was set to 'End Effector', and the command frame was set to 'External'.

The scenarios in this experiment were fly-to tasks designed to involve three-segment trajectories: either 'up – across table – down', or 'out – along table – in'. The second segment of the trajectory always involved movement along two axes (x and y, or x and z), thus requiring multi-axis hand control.

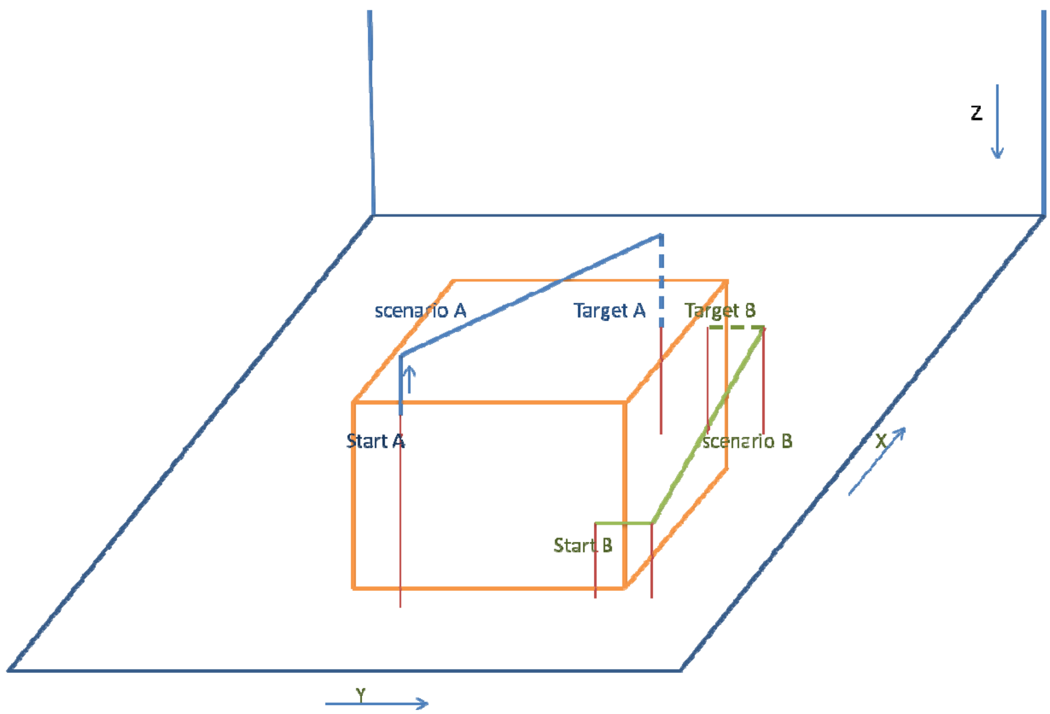


Figure 3.2: Examples of trajectory geometries. Scenario A is an example of a trajectory that goes 'up – across table – down'; scenario B is an example of a trajectory that goes 'out – along table – in'

Experimental Design

The experimental design was a 3 (trajectory control automation: auto trajectory control, trajectory guidance, and manual – between-subjects)*2 (hazard avoidance: automatic stop vs. notification only – between-subjects)*2 (camera selection automation: recommendation vs. manual selection – within-subjects) mixed design. The details of each automation function are described in Table 3.1.

The preferred trajectory involves three segments. Take the 'up-across table-down' trajectory for example, the first segment is a straight line that goes from the start point to the first waypoint that is has the same x and y coordinates but higher in position that the start point. Similarly, the third segment connects the target point with the second waypoint which is right above the target point. The second segment is a straight line that directly connects the first and second waypoints. This line is preferred because it has the shortest distance between the first and the second waypoints and at the same time maximizes the 'multi-axis hand control'. This trajectory is supposed to match the trajectory planned by the participant, who has been trained to do so. Thus the trajectory guidance (visible line and hand control input advice) is designed to provide aid for the trajectory execution, i.e. Stage 4 automation.

Table 3.1: Automation Design

	Stage 1&2: Information acquisition and integration	***Stage 3: Decision making / Action selection***	***Stage 4: Action Implementation***
	Monitoring the configuration and movement of the arm	*Selecting camera views*	*Operating (moving) the arm*
High auto	Automation stops the arm and highlights the joint in which a singularity, joint limit, or collision is about to occur.	--	Automatic EE movement/control (AC) – automation shows and flies the preferred trajectory.
Medium auto	Automation notifies the operator, using visual highlighting, if a joint is about to reach a singularity or joint limit (within 5 degrees), or if a collision is about to occur (within 0.8m from a structure, i.e. 0.2m from violating the 0.6m flight rule).	Automation suggests (changes to) camera view selection. Human makes the selection.	Automation provides guidance (AG) - automation shows a preferred trajectory within the camera views, and suggests the initial hand control input along the six degrees of freedom. Human implements the trajectory manually.

Manual	--	Human operator selects camera views and makes necessary adjustments when deemed necessary.	Human manually specifies and implements the trajectory.

The dependent measures included performance outcome measures, process measures and subjective ratings. The performance outcome measures were:

(1) Task completion time and accuracy (measured in terms of mean deviation from the desired path and from the target location).

(2) Hazard occurrence (whether a hazard, such as violation of flight rule or collision with structures, occurred).

(3) Response time to warnings (time between initiation of the warning and operator's first change to control inputs).

(4) Camera selection (number of camera switches per scenario, and percentage of time operator followed camera recommendations).

The process measures include operator's scan pattern and effectiveness (i.e., use of a hub-and-spoke scan pattern, as suggested during training, and scanning of all GUIs) which were examined using eye tracking data (fixations and percentage dwell time in areas of interest (AOI)). The AOIs include the camera views and the window view, the digital map area, the mode selection area, the brake area, the joint angle area, and the camera selection area. Eye tracking data were collected using the ASL Eye-Trac 6D eye tracking system (sampling rate 60Hz).

Other dependent measures included subjective workload ratings (mental and physical workload, on a 0-10 scale), which were measured after completion of each scenario; and perceived advantages and disadvantages of the automated aids, gathered by means of a debriefing questionnaire.

Procedure

The participants first completed a training session (lasting between 3-5 hours, depending on the participant's learning rate and their group assignment), followed by an experiment session of 3-4 hours. In the training session, participants completed two scenarios without any automated aids, two scenarios with hazard alerts and camera selection automation, one scenario with trajectory guidance and one scenario with auto trajectory control. The participants in the trajectory guidance and auto trajectory control group also completed a seventh scenario, in which they encountered hazard alerts with the trajectory automation, i.e., an automation failure where the automation proposed or flew a path that put the arm too close to a structure. All participants were able to follow the correct procedure to perform the scenarios, and were able to manually fly the arm to the target point within tolerance at the end of the fourth training scenario.

Six (manual trajectory group) or seven (all other groups) scenarios were performed during the experiment session (see Table 3.0). The first, third, fourth and fifth scenarios for all groups were routine scenarios. All three automation aids worked reliably, and no hazards were designed into these scenarios. The first and fifth scenario were completed without camera recommendations; the third and fourth with camera recommendations. The second scenario involved a designed-in hazard (the trajectory guidance and control provided a path that penetrated an alert zone) with the hazard avoidance automation alerting the participant. The seventh scenario, which used a mirrored trajectory geometry to the one in the second scenario, also included an induced hazard but in this case, the hazard avoidance automation failed and did not provide an alert. The sixth scenario (trajectory guidance and auto control groups only) involved a trajectory control failure, in which a wrong path that ended 2-3m from the actual target point was presented or flown. Table 3.2 shows the sequence and nature of all scenarios. Each scenario was followed by a short questionnaire asking participants to rate their workload and performance, and to report any warnings they noticed and the assumed underlying problem.

Following the completion of all experimental scenarios, participants were debriefed regarding their strategies for completing different components of the scenarios, any difficulties with performing the tasks, their use of and comments on automation aids, and comments on the training and experiment setup. A complete list of all debriefing questions can be found in Appendix VIII.

Table 3.2: Sequence of experimental conditions in Trials 1-7

Sequence	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6 (for AC and AG only)	Trial 7 (Trial 6 for manual group)
Condition	Good trajectory no camera recommendations	Collision encounter (with alerting success, HW or HS) No camera recommendations	Good trajectory w/ camera recommendations	Good trajectory w/ camera recommendations	Good trajectory No camera recommendations	Trajectory guidance failure or trajectory control failure No camera recommendations	Collision encounter with alerting failure No camera recommendations
Trajectory Geometry	One of Geometries 1-5	Geometry 7	One of Geometries 1-5	One of Geometries 1-5	One of Geometries 1-5	One of Geometries 1-5, with the wrong path	Geometry 6
Notes		Mirror of Geometry 6. Has a designed-in hazard. In this trial, the hazard alerting works perfectly.				This trial does NOT have a designed-in hazard encounter	Mirror of Geometry 7, around the table. Has a designed –in hazard. In this trial, the hazard alerting fails.

Results

The performance data (i.e., completion time and trajectory deviation) were inspected for normality and, if necessary, log transformations were applied prior to their analysis. One way ANOVA or repeated measures linear models (using the General Linear Model formulation in SPSS 20) were used to analyze the main effects of trajectory control automation, hazard avoidance automation, and camera recommendation automation. Tukey's test was employed for post-hoc analyses of significant findings. Data from trial 4 (T4) and trial 5 (T5) for routine performance were used, because they best typified the more skilled operator. Since interactions between different types of automation (specifically, camera recommendation and trajectory control automation) were not of interest, the data were averaged over T4 and T5 to examine the effects of trajectory automation. Data from T4 and T5 were compared to examine the effects of camera recommendation. Data from these two trials for routine performance were used, because they best typified the more skilled operator. Each trial was also segregated in to five phases: segment 1, decision time 1, segment 2, decision time 2, and segment 3. A decision time is defined as the period of time between two segments where all hand control inputs were zero as the operator tried to decide how to proceed.

Routine Performance (In the Absence of Automation Failures)

Completion Time. The level of trajectory control automation was found to significantly affect scenario completion time ($F(2, 33) = 14.46$, $p<.01$), such that participants in the manual condition took more time ($M = 442.46$, $SD=107.74$) than those in the auto guidance (AG) condition ($M = 401.58$, $SD=133.80$) which, in turn, were much slower than participants in the auto control (AC) group ($M = 215.21$, $SD=83.76$; see Figure 3.3).

Post-hoc analyses using the Tukey's test showed that completion time in the AC condition was significantly shorter than in the manual ($p<0.001$) and AG ($p=0.001$) conditions. The Manual-AG difference, however, was not significant ($p=0.64$).

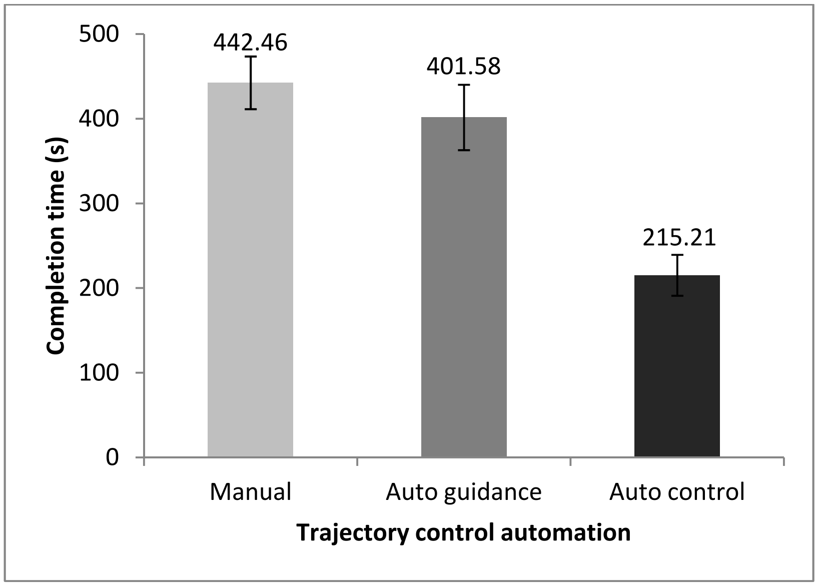


Figure 3.3: Completion time for Manual, AG, and AC conditions. Error bars represent standard error

When the completion time data were broken down on a segment-by-segment basis, a comparison of the manual and AG conditions revealed that the only significant difference occurred at the second decision point, preceding the cross-table trajectory segment, where participants in the manual condition took significantly longer before resuming movement of the arm (Manual: M = 20.17, SD=19.50; AG: M = 10.41, SD=10.07). Comparison between Trial 4 and Trial 5 showed that completion time was not affected significantly by the presence of camera automation.

Trajectory Deviations. Trajectory deviations were calculated from the integrated (over segment) average of the absolute distance between the actual EE position and the preferred path from the X, Y and Z axes. At any given time point (t), this distance was defined as the shortest distance between EE and all the points (P_i) on the path.

$$Distance(EE_t, preferred\ path) = Min \left\{ \begin{array}{c} \sqrt{\left[(EE_{x-t} - PX_0)^2 + \left(EE_{y-t} - PY_0\right)^2 + (EE_{z-t} - PZ_0)^2\right]}, \\ \sqrt{\left[(EE_{x-t} - PX_1)^2 + \left(EE_{y-t} - PY_1\right)^2 + (EE_{z-t} - PZ_1)^2\right]}, \\ \vdots \\ \sqrt{\left[(EE_{x-t} - PX_n)^2 + \left(EE_{y-t} - PY_n\right)^2 + (EE_{z-t} - PZ_n)^2\right]} \end{array} \right\}$$

This deviation reflected the operator's ability to fly the arm efficiently by using the ideal ratio between hand control inputs along different axes. For example, during the second segment, hand control inputs along the X and Y axes should be proportional to the distance along these two axes between the first and second waypoint. In the auto control condition, this deviation represents the joint-system performance, of which the automation performance is the major component. A one-way ANOVA showed a significant effect of the level of trajectory automation on overall trajectory deviation (F (2, 33)=58.66, p<0.001), and on trajectory deviations on the three segments (Segment 1: F (2, 33)=9.8, p<0.001; Segment 2: F(2, 33)=32.09, p <0.001; F (2, 33)=35.39, p <0.001; see Figure 3.4). Overall trajectory deviation was smallest in the AC condition (M= 1.37, SD = 0.49), at an intermediate level with auto guidance (M = 14.73, SD=6.52), and worst in the manual condition (M = 81.22, SD=32.86; p <0.001 for both comparison with AC and with AG). Note that the trajectory deviation in the AC condition is negligibly small, although not exactly 'zero', due to limitations of the BORIS simulation.

When the trajectory deviation was analyzed separately by segment, the first segment (1-axis movement, along negative z direction) showed a benefit for AC over the other two less automated conditions, which did not differ significantly. The deviations during Segment 1 under manual and guidance conditions were less than 6cm which can be explained by the fact that the operation involved hand control along only one axis – moving up along the -z direction. The second segment (2-axis cross-table, along both x and y directions) showed the largest trajectory deviation overall. Performance was better in the two automation conditions (AG and AC which, again, did not differ significantly) than in the manual condition. A similar pattern was observed for Segment 3, in part because this segment inherited the deviations at the end of Segment 2. Table 3.3 shows the mean trajectory deviations during each segment under the three conditions (manual, AG and AC) and the results from the statistical analyses.

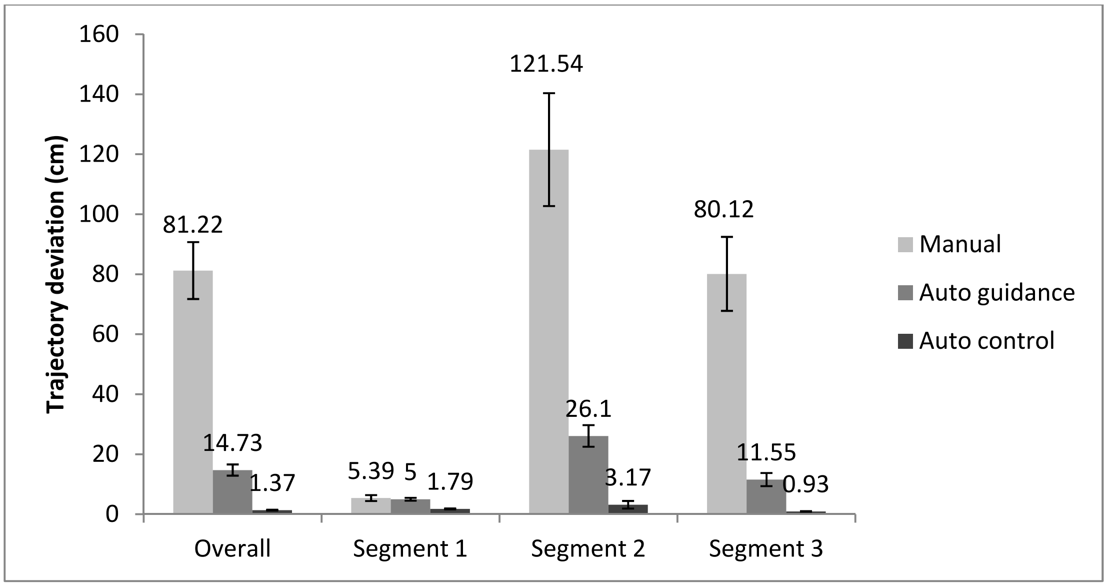


Figure 3.4: Trajectory deviation of entire trial, Segment 1, Segment 2, and Segment 3 under different levels of trajectory automation

Table 3.3: Trajectory deviation by segment

	Manual		Auto Guidance		Auto Control		Main effect		Post-hoc		
	M	SD	M	SD	M	SD	F	p	Manual vs AG	Manual vs AC	AG vs AC
Segment 1	5.39	3.39	5	1.6	1.79	0.55	9.8	<.001	0.889	0.001	0.003
Segment 2	121.54	65.16	26.1	12.52	3.17	4.39	32.09	<.001	<.001	<.001	0.322
Segment3	80.12	42.68	11.55	7.6	0.93	0.28	35.39	<.001	<.001	<.001	0.558

A 2 (trajectory automation level: manual vs. AG) X 3 (camera automation: present vs. absent) repeated measures ANOVA was conducted on the log-transformed overall trajectory deviation data. The analysis revealed a significant benefit for trajectory guidance (F (1) = 89.23; p <0.001) and for camera automation (F (1, 22) = 4.53; p =0.045), with no significant interaction between the two variables. Data for the manual and AG conditions are shown in Figure 3.5 below.

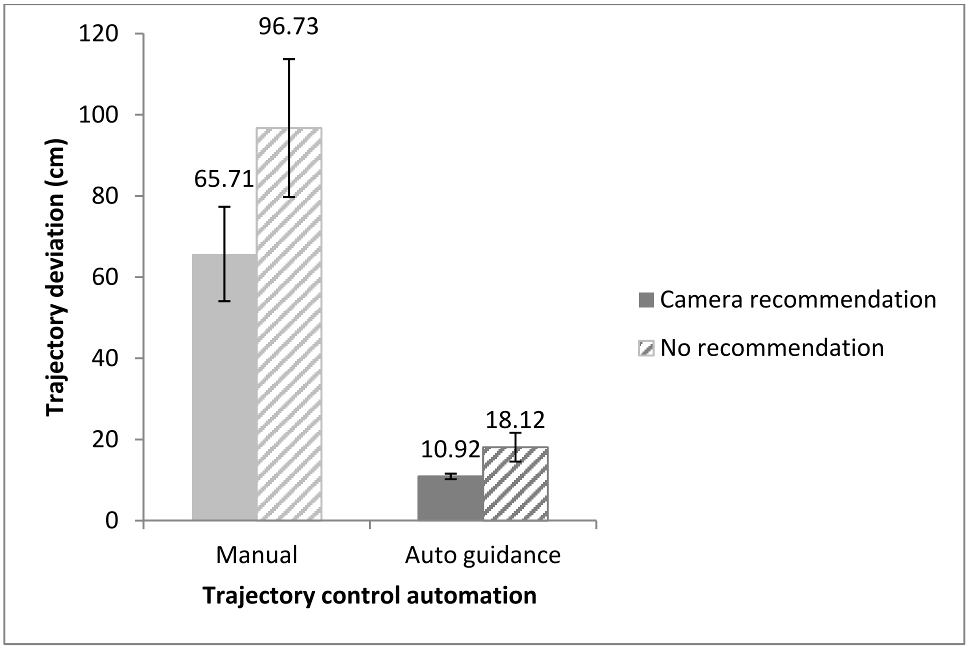


Figure 3.5: Trajectory deviation for manual and AG conditions, separated by with and without camera recommendation. Error bars represent standard error

Camera Switches. A 2 (camera recommendation: with automation vs. no automation) x 3 (trajectory automation: manual, AG, AC) repeated measures ANOVA showed that the number of times subjects changed camera views varied significantly across trajectory automation levels (F (1, 33) = 7.29; p =0.011). Participants in the manual trajectory control condition switched most frequently (M = 15.21, SD=12.19), followed by the AG group (M = 10.88, SD=6.05). Participants in the AC condition switched least often (M = 6.96, SD=2.68). There were also more camera switches when camera automation was activated (trial 4, M = 12.8, SD=10.52) than when it was not (trial 5, M = 9.1, SD=5.66; F (2) = 4.62; p<.02). While the interaction between the two variables was not significant (p =0.234), the direction of effects suggests that the main increase in camera switches with camera automation occurred in the manual trajectory condition. Figure 3.6 shows the average number of camera switches as a function of camera and trajectory automation.

Among these switches, the percentage of 'follows', i.e. the camera's switches that complied with recommendations (or the hypothetical recommendations in trial 5) as opposed to a switch to some other camera, did not differ significantly between the two camera recommendation conditions. In trial 4 (with camera recommendation) the mean follow-to-switch ratio was 64.80% (SD=14.10), and in trial 5 (no recommendation) the ratio was 61.15% (SD=17.32). Therefore, the camera recommendation did not necessarily lead to better camera selections (using the recommended camera as the judging criterion).

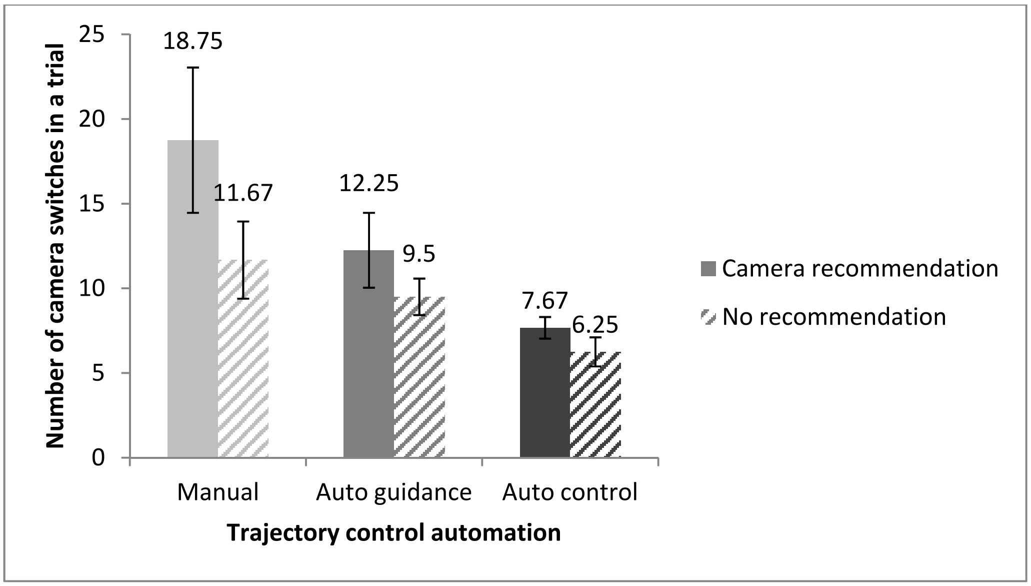


Figure 3.6: Average number of camera switches as a function of camera and trajectory automation. Error bar represents standard error

Workload. A similar pattern of results to those described above was observed for overall subjective workload (see Figure 3.7), which was rated by participants at the end of each trial. Ratings were lowest in the AC condition (M = 2.35, SD=0.95), followed by AG (M = 3.41, SD=1.09) and highest in manual (M = 4.63, SD=1.47). Paired contrasts between adjacent levels of trajectory DOA were all significant ($p<.05$) and hence, as with trajectory error, both guidance (AG) and control (AC) automation facilitated performance. Camera automation did not affect workload.

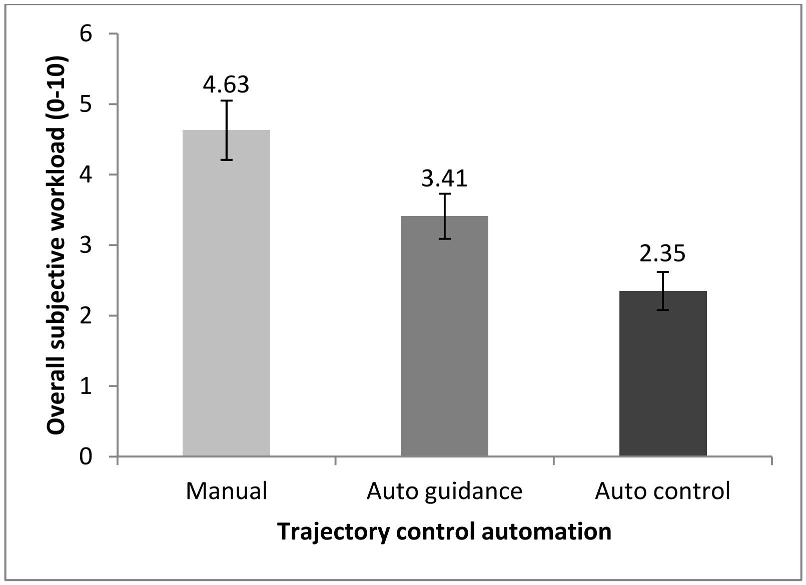


Figure 3.7: Overall subjective workload as a function of trajectory automation level

Scan pattern. Percentage dwell time (PDT) in eight areas of interest (AOIs, see Figure 3.8) was examined across the three trajectory automation conditions.

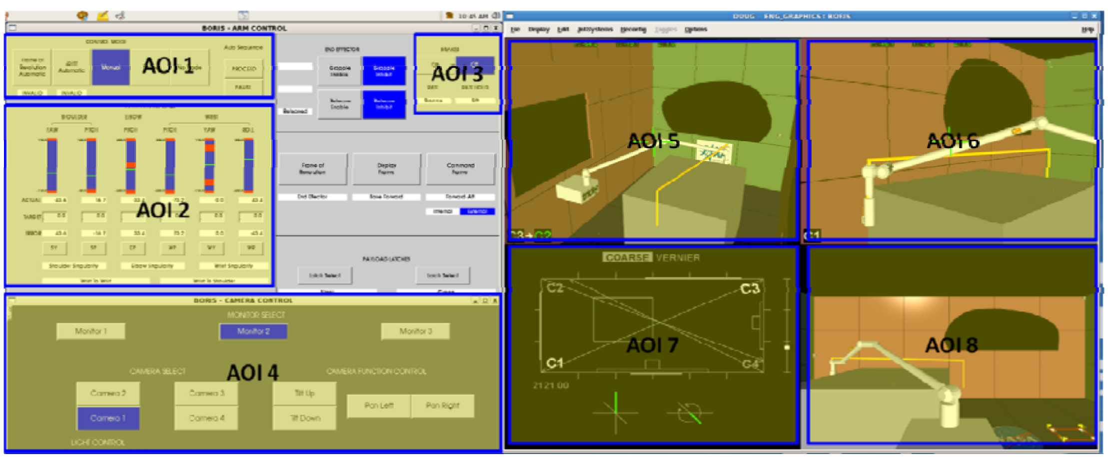


Figure 3.8: Areas of interest (AOI). AOI1: mode selection display; AOI2: arm configuration display; AOI3: brakes and rate selection; AOI4: camera selection display; AOI 5: camera view 1; AOI 6: camera view 2; AOI 7: digital map; AOI 8: window view.

A one-way ANOVA found that the level of trajectory automation significantly affected PDT in AOI 2 - Arm Configuration Display ($F(2, 33)=3.388$, $p=0.049$). Participants scanned this AOI more frequently when performing the task with AC

(PDT M=3.06, SD=3.54) than when performing the task manually (M=0.23, SD=0.27, p=0.04). The difference between the other pairs of conditions was not significant.

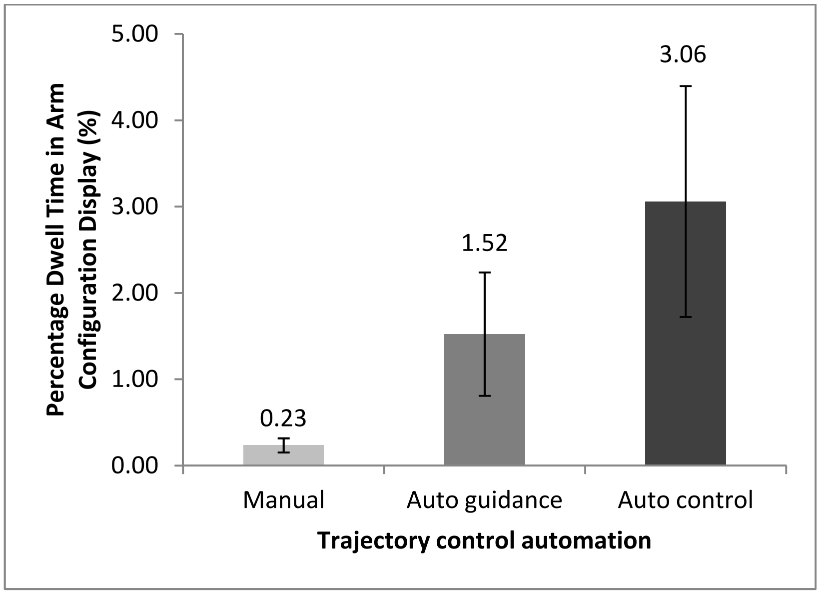


Figure 3.9: Percentage Dwell Time (PDT) in the Arm Configuration Display AOI. Error bars represent standard error

Performance In Case of Automation Failures

Trial 2 – Trajectory guidance failure

On trial 2, the trajectory guidance (shown as three line segments in the camera view for both the AG and AC group, and flown by the automation in the AC group) failed, bringing the EE within 0.8 meter of the table (close to a violation of the flight rule – 'no maneuver within 0.6 of structures'), and hence activating either the warning light (for half the subjects) or the auto-stop function (for the other half). The number of instances a participant failed to respond to a warning, the number of actual collisions and the number of non-collision violations of the 0.6m limit on trial 2 are shown in Table 3.4. Since the level of automation for hazard avoidance (auto-alert vs. autostop) did not significantly affect performance, the entries in Table 3.4 are collapsed over both groups.

Table 3.4: Collisions and violations as a function of trajectory automation level in Trial 2

	Manual (N=12)	Auto guidance (N=12)	Auto control (N=12)
# No response	0	0	2
# Violations	0	1	4
# Collisions	0	0	0

A chi-squared analysis of the 'no response' data revealed no significant difference between the three groups (although there is a trend favoring safer behavior with less automation of trajectory control). For violations, there was a marginally significant difference between the groups (χ^2 (2) = 5.7; p=0.057), clearly penalizing participants in the auto control (AC) condition, compared to those flying manually.

Trial 6 - Trajectory guidance failure

On trial 6, for both automation groups, the proposed trajectory led the end effector to an incorrect coordinate at the second waypoint, and subsequently an incorrect end point that was right below the second waypoint. In both groups, only a small number of participants (AG: 2 of 12; AC: 3 of 12) failed to notice the problem. Success or failure in noticing the problem was determined based on participants' verbalizations during the scenario (e.g., a participant might say: 'camera 3 view doesn't look correct ') and/or participants' adjustments of the arm trajectory after noticing the incorrect waypoint or end point. Not every participant who noticed the failure attempted to correct for it which may be explained by difficulties with determining whether it was indeed the result of an automation failure or a misperception due to the camera angle. Participants who noticed and confirmed the automation failure en route were more likely to correct it, and thus reduced overall error and improved EE accuracy at the end point. Figure 3.10 shows the number of participants who noticed the problem, attempted to correct it, and completed the task successfully in both automation conditions (N=12 for each condition). Chi-square tests did not show significant differences between the two conditions on these three measures.

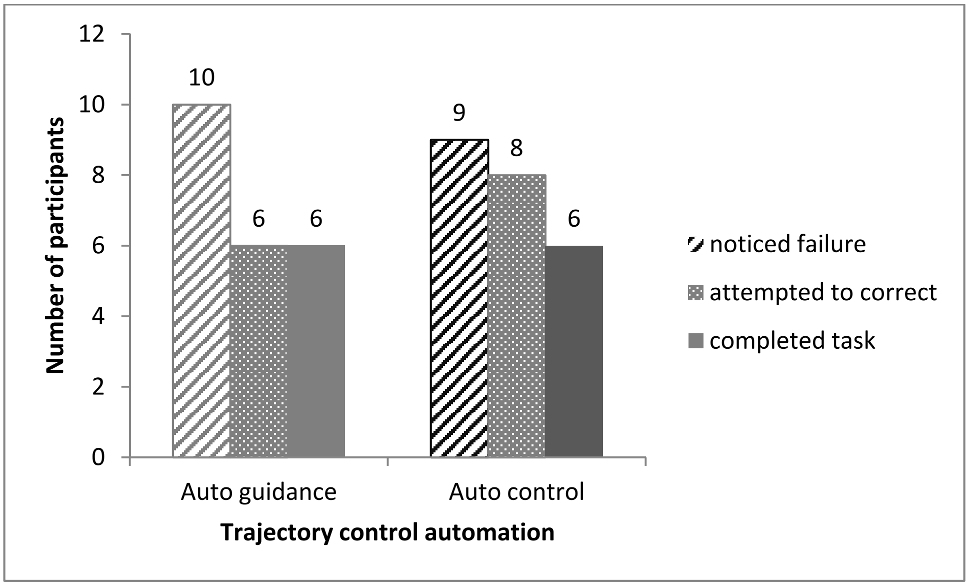


Figure 3.10: Number of participants in the two automation conditions who noticed the failure, attempted to correct the error and completed the task successfully (n=12 in each condition)

Independent samples t-test revealed that the AC group experienced significantly larger deviations (M=143.89, SD=20.29) than the AG group (M=113.35, SD=29.53; t = 2.95; p=0.007). Figure 3.11 shows the trajectory deviation in the failure trial for the AG and AC conditions and the trajectory deviation for the manual condition in the routine trials.

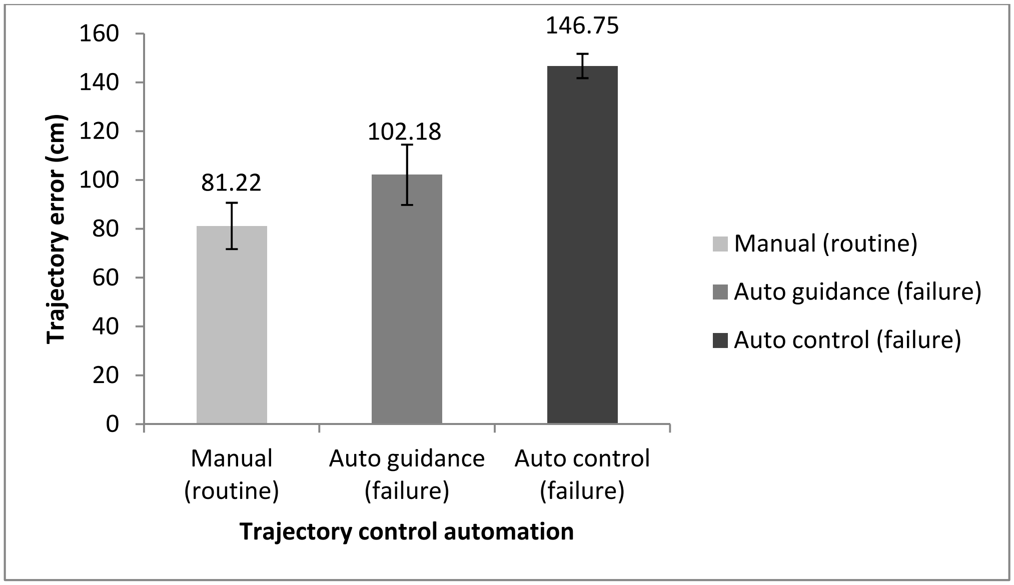


Figure 3.11: Trajectory deviations for the AG and AC groups in trials involving a trajectory automation failure (compared to trajectory deviations for participants in the manual condition during routine tasks)

Final trial - Hazard avoidance failure

On the final trial (trial 7 for the two automation groups, trial 6 for the manual group), the trajectory coordinates directed the EE into a proximity violation relative to the table. However, unbeknownst to the subjects, the collision protection system failed to alert (half of the participants) or stop the arm (the other half of the participants) when the warning boundary was crossed. On these trials, all but one participant in the manual group moved the arm into the warning zone (note that they did not receive a warning). The number of participants who did not show any response to entering a warning zone was 1 (manual), 3 (AG) and 8 (AG). The difference between the three groups was marginally significant (χ^2 (2)= 9.14; p=0.10). The number of subjects who allowed a violation to occur was 8 (manual), 5 (AG) and 12 (AC) - again a proportion that significantly differed from expectations (χ^2 (2)=10.02; p=0.007; see Table 3.5). Poorest performance was again observed for the auto-control group.

Table 3.5: Collisions and violations as a function of trajectory automation level during hazard-avoidance failure trials.

	Manual (N=11)	Auto Guidance (N=12)	Auto Control (N=12)
# No response	1	3	8
# Violations	8	5	12
# Collisions	2	0	0

In light of the differences between the manual, AG, and the AC conditions in terms of failure detection performance, differences in scan patterns were examined between these two groups on trial 5. The goal was to determine whether the poor failure detection for the AC group could be explained by complacency-induced visual neglect of the two camera views, where the effects of failures would become evident. However, there was no significant difference ($p>0.10$) between any of the three conditions with respect to the percentage dwell time on the two monitors.

Summary of responses to debriefing questions

Participants were asked to describe how they used the automated aids and comment on the benefits of these aids for various aspects of the task, including: choosing camera views, monitoring the arm configuration and clearance, monitoring the arm movement, and flying the arm.

Trajectory control automation. Twenty eight participants said the trajectory control automation was helpful for flying the arm. These included 8 participants from the manual group who had performed tasks with both levels of the trajectory control automation during training. Participants thought trajectory control automation made operation of the arm precise, smooth, and efficient. Participants found that, with the visible path, they were able to see where the arm was going, and how far it had to go. The trajectory guidance helped operators focus on monitoring hazards/clearance/and views, rather than focusing on the controller. Trajectory control automation was considered as an effective means of minimizing the chance of collision and also ‘helpful for camera selection’, because the visible line helped visualize the path, so the operator could plan the camera changes.

Hazard avoidance. Thirty-one of the 36 participants thought that hazard avoidance was helpful in monitoring the arm configuration and/or monitoring the clearance. The high level of hazard avoidance automation also “allowed the operator to make sure that the arm was not going to collide with a structure or object”. Participants also commented that the hazard avoidance “provided more room for error, [and was] thus more comfortable”.

Camera recommendation. Among the 36 participants, seventeen felt camera recommendations were helpful because they provided perfect views and could be used to cross-check the appropriateness of the participant’s choice. Camera recommendation could also be helpful for monitoring the clearance between the arm and structures (like walls and the workbench) and monitoring arm configuration. In addition it could make flying the arm easier. Four participants thought that recommendations were helpful but not necessary. For example, one participant said: ‘this was great but I did not need automation’. Seven participants explicitly said that camera recommendations were not useful at all. They preferred to choose their own camera views, or felt that the recommendation “did not tell me anything I don’t know”. The blinking of the recommendation was perceived as annoying or distracting.

Discussion

The study described in this chapter evaluated the effectiveness of three types of automation: hazard avoidance (stage 1 and 2 automation), camera recommendations (state 3 automation), and trajectory control automation (stage 4 automation). For each of the three systems, performance at one or two levels of automation was compared with a manual condition, in the context of both routine and automation failure scenarios.

Performance progressively improved as the level of trajectory control automation increased. The higher levels led to faster (as indicated by trial completion time), and more accurate and efficient (measured in terms of trajectory deviation) operation of the arm. These trends were monotonic, and adjacent pairs along the level of automation (LOA) scale were always significantly different, except for the completion time between the manual and AG conditions. Each added level of automation assisted performance in different ways. At the medium level, the presentation of the proposed trajectory in the camera views resulted in reduced trajectory deviations and improved efficiency because the line served as a visible reference, thus transforming an otherwise cognitively demanding task into a much easier perceptual tracking task. However, since the operator still had to fly the arm manually, completion time did not benefit. The high level, auto control, removed the planning and coordination requirements involved in manual control, thus improving performance further, both in terms of trajectory deviations (which was essentially zero) and completion time. The two benefits of increasing the level of trajectory control automation, i.e., reducing cognitive demands by changing the task into a perceptual one and eliminating the need to plan and execute hand control inputs, also led to reduced mental workload. Consequently, operators had a more balanced scan pattern as exemplified by the percentage dwell time in the Joint Angle Display.

While the level of camera recommendation automation was not varied, the effects of its presence or absence can be compared. Participants followed the advice some of the time - 30% more camera view switches were observed when the advice was available, compared to when it was not. This led to enhanced performance in terms of higher trajectory accuracy and efficiency (smaller trajectory deviations) since participants were getting a better view of the EE relative to the path. These benefits were seen independent of whether the arm was controlled manually or following automated guidance (AG). The camera recommendations did not affect workload or completion time, however. One possible explanation for this finding is that the camera selection task did not offer very many choices and may therefore not have added much to overall mental workload even when it was done by the participants themselves.

The debriefing results also show that participants did not consider the camera recommendations to be very useful. This may be explained, in part, by limitations of the algorithm employed to generate the camera recommendations. First, the algorithm merely considered the position and movement of the EE but not any other segment of the arm (such as wrist, boom, and elbow). As a result, the recommended camera views supported a good task view but did not necessarily provide a clearance view (i.e., a view of the distance between the arm and the FORWARD wall, for example). Second, the camera view recommendation did not give advice on the proper orientation of the camera, i.e. pan and tilt adjustments. Therefore, even if the recommended camera was selected, the view from that camera may not have been very helpful and the participant may have chosen a different camera in the end. Third, at times, the quality of the four camera views did not seem to differ significantly. In those cases, the FORT values for the cameras were very similar, and the view from any camera was satisfactory.

In addition to the limitations of the camera recommendation algorithm, participants may have under-used the recommendations for two other reasons: (1) they had planned their camera selection at the beginning of a scenario, and once the operation started, participants tended to follow their original plan, and (2) some participants had strong camera preferences (e.g., some participants preferred to use camera 3 (the one on the ceiling) as the 'big picture' view, independent of the specific task).

Finally, the level of hazard avoidance automation was varied, from one merely alerting to potential hazards to one that prevented them. This increase in the level of automation did not have any effect on performance which can be explained by a 'ceiling effect'. Participants responded to entering a warning zone nearly 100% of the time, both in the alert only and in the alert+stop conditions. Thus, an alert alone was sufficient for preventing the arm from moving closer towards a structure or an object.

Similarly, the effectiveness of warnings alone was observed in the experiment described in Chapter 2.

Overall, across the three types of systems, increasing levels of automation led to improved performance and lower workload during routine scenarios. However, in case of automation failures, the expected performance costs (Bainbridge, 1983; Perrow, 1986; Reason, 1997; Woods, 1996) at high LOAs were observed. They can be explained by automation-induced complacency (Moray & Inagaki, 2000; Parasuraman, Molloy, & Singh, 1993) which can lead to difficulties with taking over from the automation and returning to manual performance, as demonstrated in previous studies (e.g. Endsley & Kiris, 1995; Parasuraman et al., 1993; Manzey, Reichenbach, & Onnasch, 2012). For the high level of trajectory control automation (AC), these problems were evident. On all three failure trials, participants in this group performed substantially worse in terms of failure recovery, than did the auto-guidance group which was required to remain in the control loop. Particularly in trial 6 (AG and AC presented and flew a path that led to an incorrect target point), the performance cost was considerable. Participants failed to notice the problem, or failed to correct for the automation failure, and consequently 50% of them failed to complete the task. Also, completion time and trajectory accuracy suffered. In the failure scenario, trajectory deviations with AC were larger than for AG. Both of them were larger yet than the manual trajectory deviation in a routine scenario. The completion time under AG and AC in the failure scenario were not significantly different from each other; however, they were both as long as (if not longer than) the manual completion time in a routine scenario (see Figure 3.12 for performance data on both routine (trial 4&5) and trajectory automation failure (trial 6) scenarios).

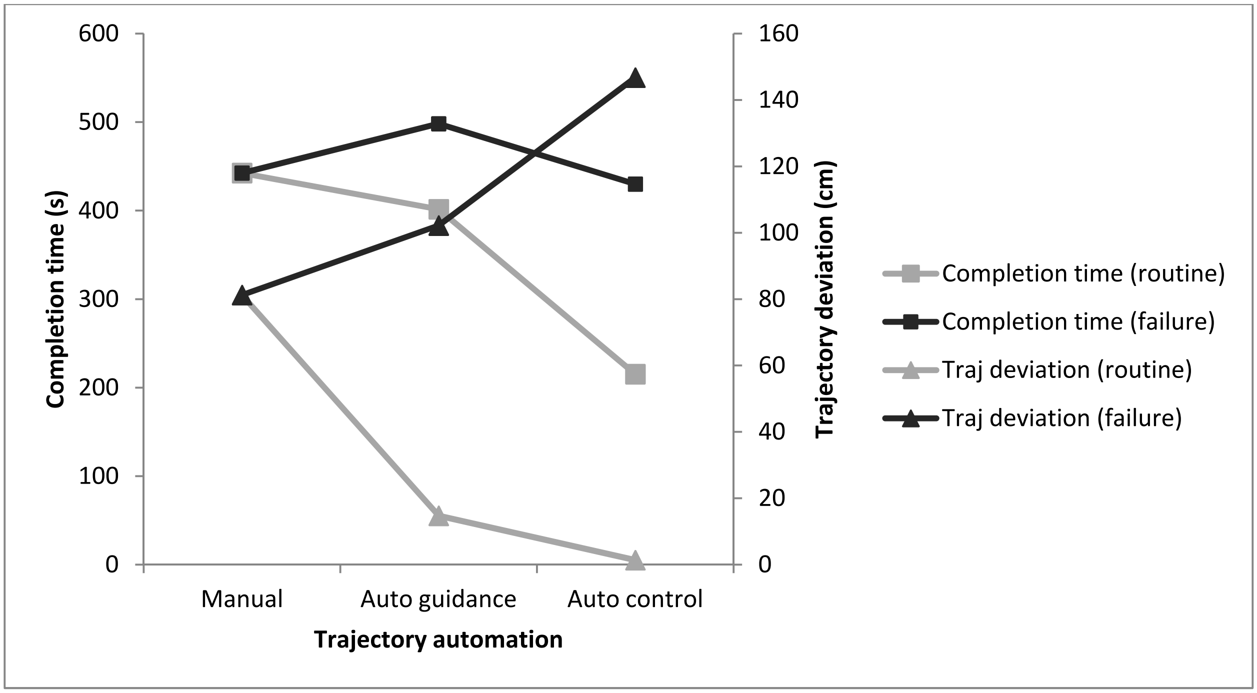


Figure 3.12: Effects of level of trajectory control automation on performance: comparison between routine and failure scenarios

To explain the performance costs during failure trials, we need to examine participants' responses to noticing the problem. We can divide participants into two groups. The first group believed (incorrectly) that the EE was in the correct target location. A small number of participants in this group did not check the EE position, as they were supposed to do, at the completion of the trajectory. This suggests overreliance on (see Lee & Moray, 1994; Lee & See, 2004; Lerch, Prietula, & Kulik, 1997; Madhavan& Wiegmann, 2004; Masalonis & Parasuraman 2003; Mosier, et al., 1992; Muir & Moray, 1996) and misuse of (Parasuraman & Riley, 1997) the automation. Another group of participants checked the EE position, but believed that it was correct either because they misinterpreted camera views, or because they paid attention only to those camera views that suggested proper positioning of the EE. This suggests that they may have been subject to confirmation bias, in that they sought out only confirming evidence (camera views) while ignoring contradictory information (Wickens & Hollands, 2000). It is important to note that this behavior occurred primarily towards the end of scenarios, when the participants experienced fatigue and higher mental workload (Cook & Woods, 1994; Janis, 1982; Sheridan, 1981; Wright, 1974).

The second group suspected that the EE was in the wrong location and attempted to correct the problem. The increase in completion time can be attributed, in part, to the participants' extensive efforts to diagnose the problem. Upon noticing a deviation between the target and the actual EE position, the participants tended to check all four camera views multiple times before concluding that an automation failure had indeed occurred. They struggled because, on the one hand, they saw that the EE was perfectly aligned with the target point on the proposed trajectory; on the other hand, the EE was not in a location where it aligned with the markers on the walls. When they tried to resolve the problem, the EE was no longer aligned with the path. These two conflicting pieces of information caused participants to hesitate because they were not sure who to trust, the automation or themselves. One participant resolved this conflict by going 'half and half'. In other words, she moved the arm to a location that was in the middle between the incorrect automation-chosen destination and what she considered to be the correct target point.

Participants in the AG and AC groups also tended to miss the opportunity for early detection of the problem. They did not check the accuracy of the proposed path before the arm was moved (manually or automatically). This failure to check the path and subsequently follow an incorrect trajectory represents an example of a combined omission and commission error due to automation complacency (Parasuraman, et al., 1993; Moray & Inagaki, 2000; Sheridan & Parasuraman, 2006) and 'automation bias' (Mosier & Skitka, 1996), where the availability of an automated aid leads a user to ignore available information. In the current experiment, the participants failed to examine the alignment between the end point of the path and the markers on the walls, and they did not compare the displayed path with the drawing on the map. Instead, they blindly followed the automation advice.

Overall, the findings from this study support the 'lumber jack' hypothesis and replicate similar findings from previous research. For example, Sarter and Schroeder

(2001) examined the performance of pilots interacting with an automated decision aid that supported decision making in case of in-flight icing events. Two types of decision aids were compared: one – a status display - provided information about the specific icing condition only (which implies required responses and pilot actions); the other – a command display – told pilots how to respond to the icing encounter. Performance benefit (increased number of correct decisions and responses) was observed when the aids provided correct recommendations. In case of inaccurate information, however, the aids resulted in performance decrements (e.g., a 100% increase in stall rate) compared with the baseline condition (no decision aid). As in the present study, this effect was stronger for the more highly automated aid (the command display).

In summary, the three types of automation in this study were overall successful in supporting operators in space teleoperation tasks. The automated aids improved operator performance and system safety, and reduced subjective workload in routine scenarios. The benefits were more pronounced as the level of automation increased. However, the costs of unreliable automation during failure scenarios also increased with LOA. A medium level of automation, particularly for the trajectory control automation, should be considered in future system design as it provides reasonable support during routine operations but helps avoid potentially catastrophic outcomes in those rare circumstances when automation fails. The findings from this experiment lay the foundation for the design of dynamic function allocation schemes in Chapter 5.

Chapter 4

Stages and Levels of Automation: A Meta-Analysis

Introduction

The findings reported in the previous chapter revealed a tradeoff between the benefits of high levels of automation during routine scenarios and the costs of such automation in case of automation failures. Similar results have been obtained in earlier studies (e.g. Endsley & Kiris, 1995; Sarter & Schroeder, 2001) that varied either stage or level of automation and examined the effects of degree of automation (DOA). Figure 4.1 shows the tradeoff between routine and off-nominal performance, as well as a hypothetical tradeoff between two secondary variables, workload and loss of situation awareness.

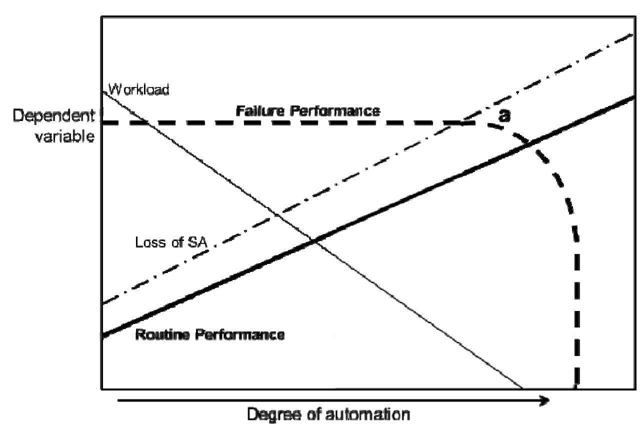


Figure 4.1: Trade-off of variables, with degree of automation (adapted from Wickens, et al. 2010)

However, few efforts have been made to integrate these findings to confirm and analyze the tradeoff in more depth. To fill this gap, a meta-analysis was conducted that (1) aggregates data from previous studies on the effect of DOA , (2) examines the extent to which these studies confirm the tradeoff between high level automation effects during normal operations versus failure conditions, and (3) identifies factors that may mitigate or moderate this tradeoff.

Method

To identify studies to be included in this analysis, we performed literature searches using scientific literature databanks (e.g. Google Scholar), reviewed tables of contents of relevant journals (e.g. Human factors, Ergonomics, International Journal of Human-Computer Interaction) and read relevant conference proceedings from the past two decades. Only studies that compared at least two different degrees of automation, either by varying the stage of automation or by varying the level of automation within a stage, with respect to at least one relevant performance measure were included. Ultimately, the data from 17 studies in a variety of domains (e.g., process control, aviation) were combined for this meta-analysis (Rosenthal 1991; Fadden, Ververs, & Wickens, 1998; Horrey & Wickens, 2006). Different from earlier analyses, effect size measures (which were not always reported) were not aggregated. Instead, a coarse three-level scale of "effect" was employed, which is defined as "non-significant" ($p>0.10$), "marginally significant ($0.10>p>0.05$), and "significant" ($p<0.05$). A similar approach was applied by Wickens and Dixon (2007) whose study examined stage-2 automation reliability.

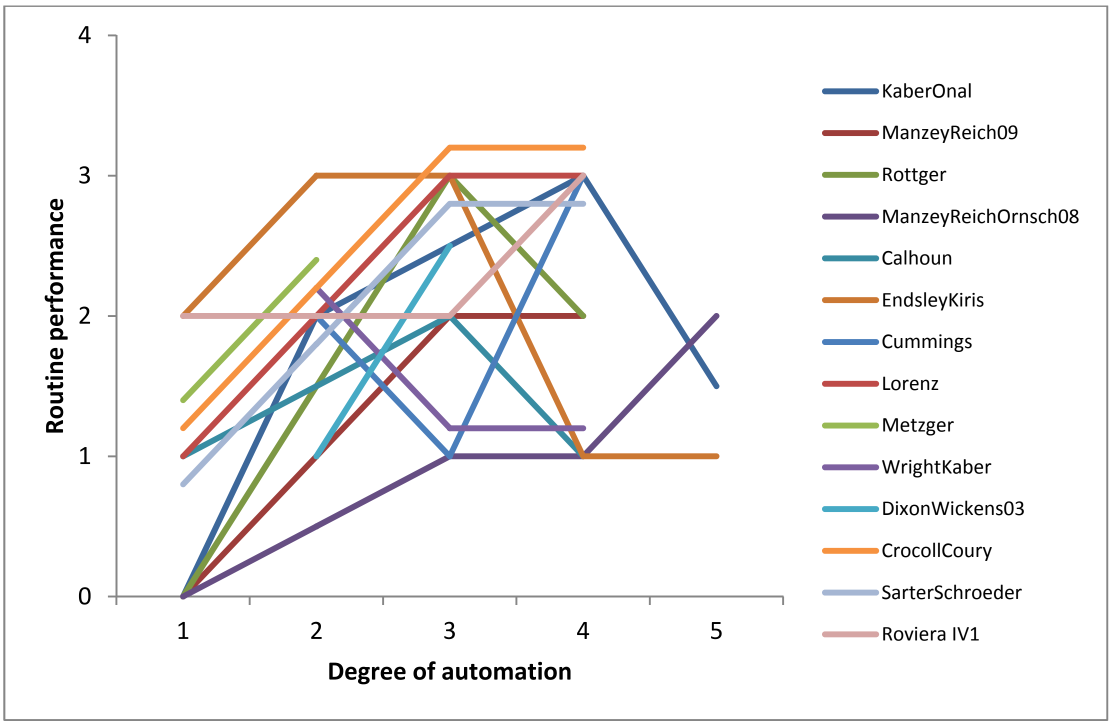
For those studies where performance was assessed by more than one dependent variable (e.g., speed and accuracy), we combined those measures to arrive at our "effect" measure. For example, if speed favored one condition in a comparison, but accuracy favored the other by an amount within the same statistical category, it was considered as a "null" effect. This only happened in one of the studies. The "effect" of interest was the effect of increasing the degree of automation.

Two independent coders determined for each study (1) the automation conditions it compared, (2) whether or not they included a manual baseline, and (3) what stages/levels of automation defined the experimental conditions. We then coded the magnitude of performance effects (non-significant, marginally significant $.10>p>.05$, or significant, $p<.05$) as a function of increasing degree of automation for as many of the four dependent variables of interest as possible: routine and failure performance, workload, and situation awareness. If a study examined three conditions and showed a significant omnibus F test, but a post-hoc test showed only one significant contrast (for example, between the lowest and highest score), we classified the latter two contrasts as "marginal". This strategy of coding the performance effects eliminated the need to normalize different types of data (e.g., to compare performance across varying workload or situation awareness measurement techniques).

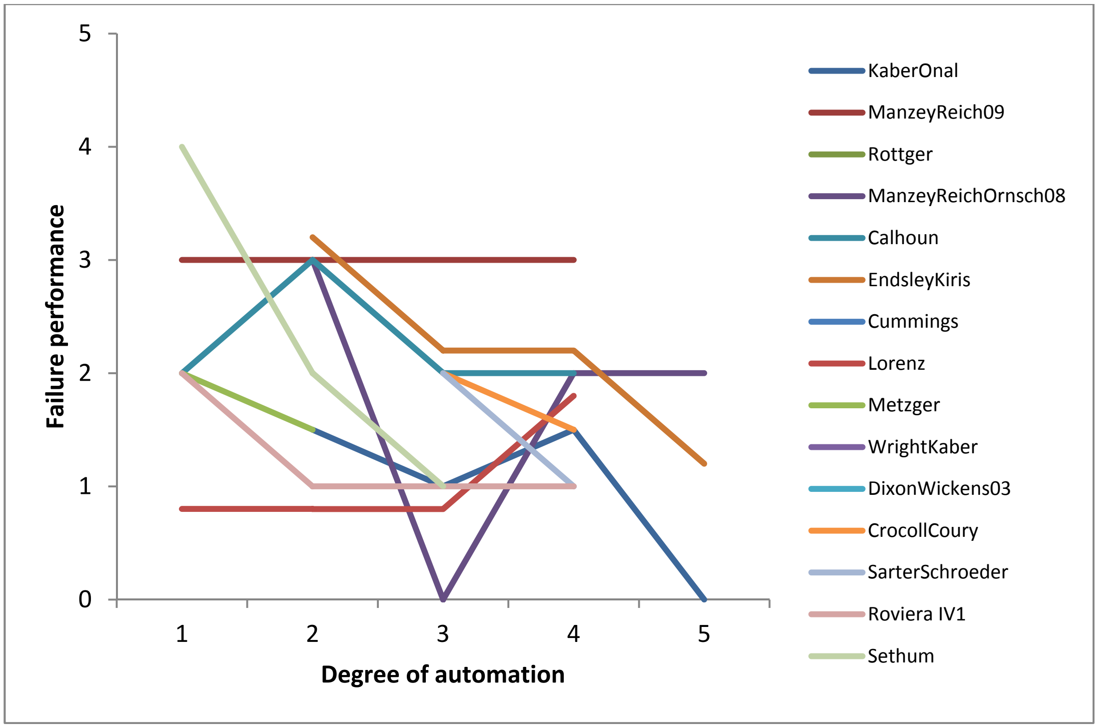
In the current analysis, all but one of the studies varied the stage of automation, either by itself while maintaining the same level of automation (e.g., high stage 2 vs. high stage 3) or, more frequently, by adding a later stage and combining it with an earlier one (e.g., stage 2, less automation vs. stages 2&3, more automation). The one exception was a study that kept the stage constant (stage 3) and varied level only at that stage (Rovira, McGarry & Parasuraman, 2007).

Results

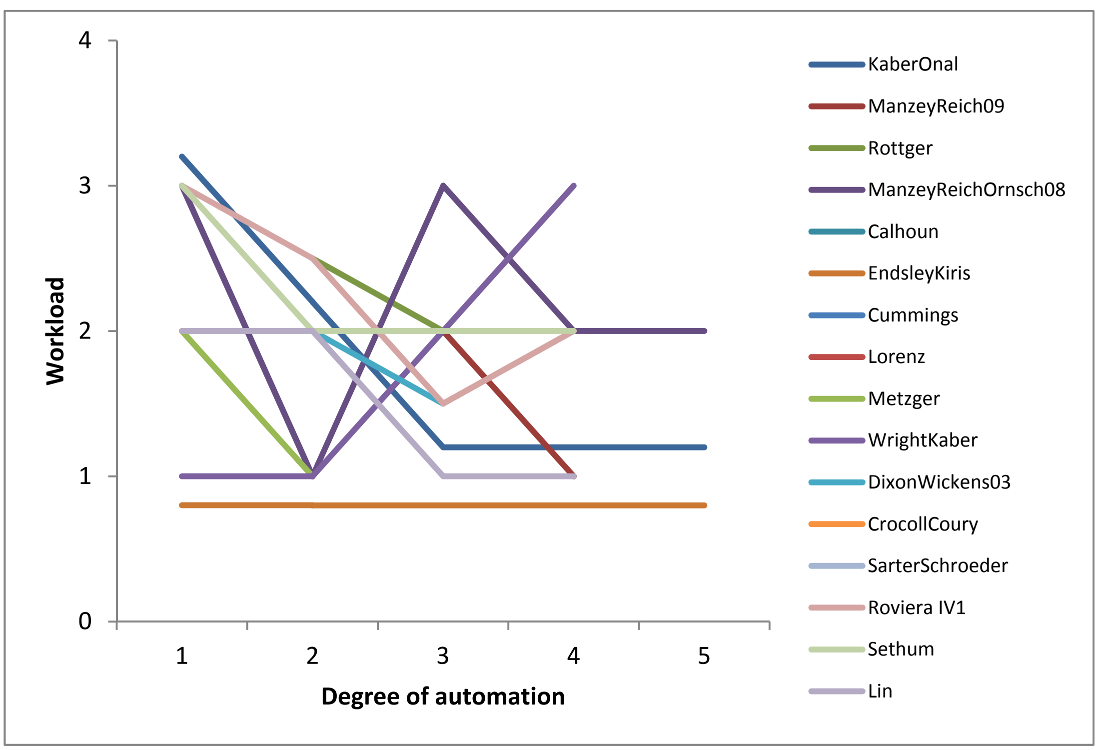
Figures 4.2 A-D show the aggregated data for each of the four dependent variables. The goal of these figures is to capture and visualize general trends (or lack thereof) across studies. A high score on the Y axis in each plot represents better routine performance, better performance under failure, higher workload (note here that higher is "worse") and higher situation awareness, respectively. Each line connects the conditions included in a given study, ordered by increasing degree of automation from left to right. If a baseline is included, this is always the left-most point indicated by an X-axis value of 1. Note that in this rendering, corresponding non-zero X axis values do not always represent the same stage/level, since these degrees of automation are presented ordinally instead of absolutely within each study. Note that three of the studies in the failure performance graph (see Figure 4.2 B) include data points in the baseline condition (0 degree of automation). In these cases, the failure was not an automation failure per se, but rather a system failure (e.g. aircraft engine failure) that participants needed to detect in a manual condition or with varying degrees of automation.



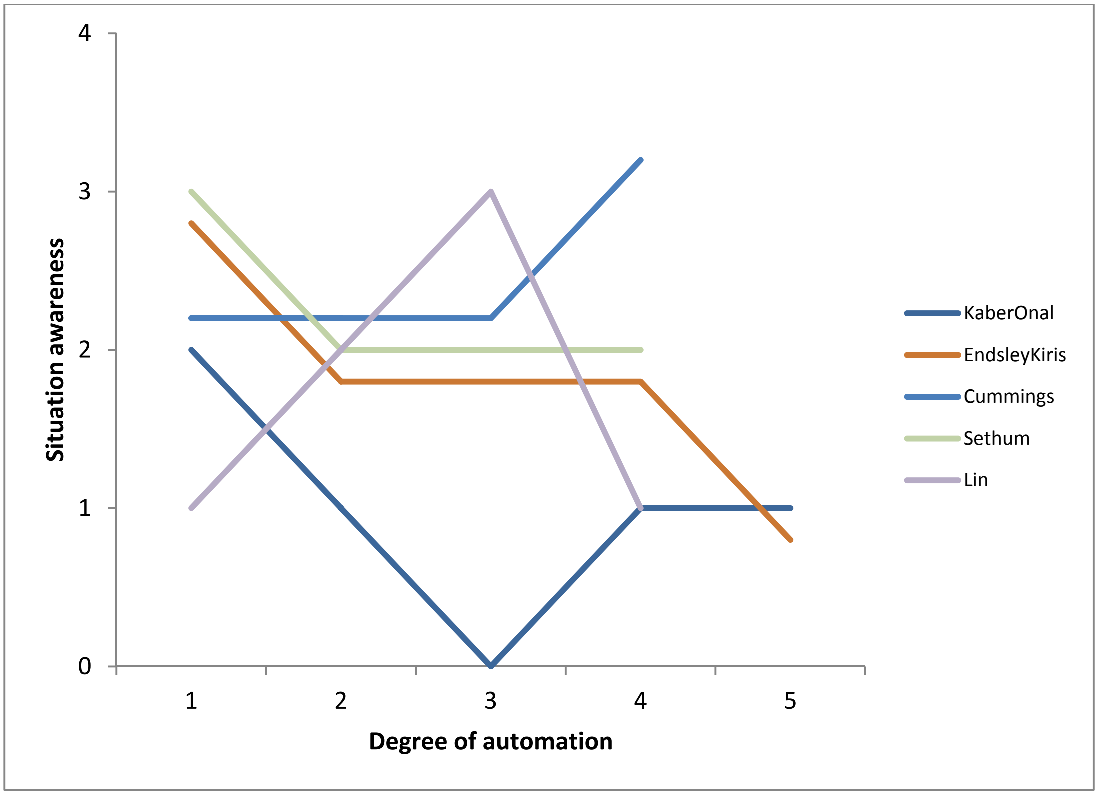
(a) Routine performance



(b) Failure performance



(c) Workload



(d) Situation awareness

Figure 4.2: Empirical data on performance, workload, and situation awareness as a function of degree of automation.

The results from each study in a given panel can be represented by a slope measure, indicating how the variable of interest changes as a function of degree of automation. For example, a steep positive slope for routine performance indicates that increasing the degree of automation produces substantial gains in routine performance. When a study is represented by 2 data points, the slope index is the slope of the line connecting the two data points. When a study is represented by 3 or more data points, this index is the slope of the regression line that fit all data points. See Figure 4.3 for an example of the regression line for one dependent measure (routine performance) in an individual study. The slope for this example was 0.7.

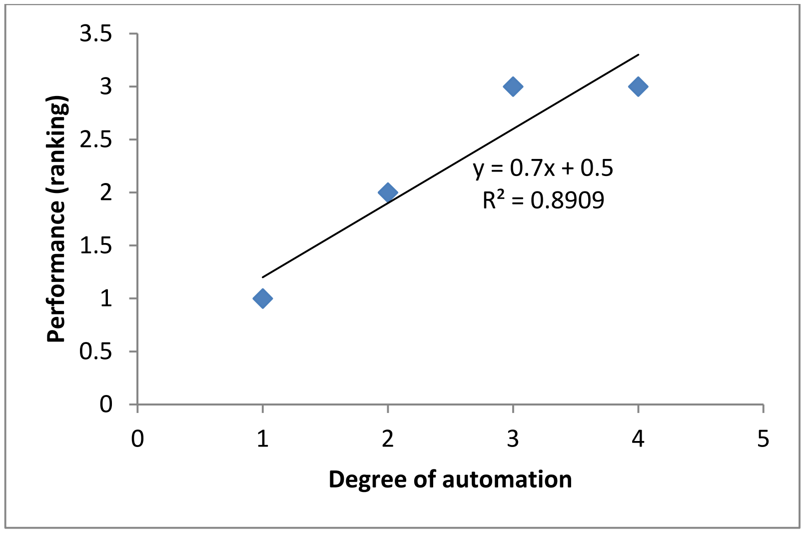


Figure 4.3: Regression line and slope for routine performance data for Lorenz et al. (2001)

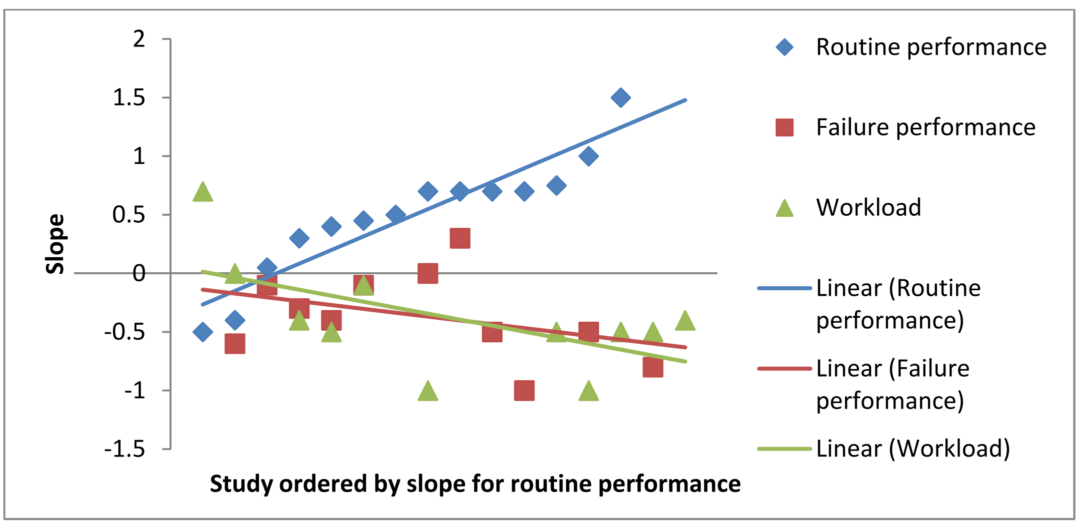


Figure 4.4: Data points arranged by increasing benefit of increasing automation to routine performance (squares, heavy regression line). Benefit is quantified by the slope of each study represented in Figure 4.1 top panel. Each X axis position plots all variable.

In Figure 4.4, the slope measure for each study is plotted as a data point on a single scale. Studies are rank ordered on the X-axis by increasing slopes (benefits for higher automation) for routine performance (blue diamonds, fitted by the blue solid regression line). The routine performance measure for each study is then connected to slope measures for the other variables of interest (failure performance, workload, and situation awareness). The green triangles indicate how workload changes as routine performance progressively benefits from increasing degrees of automation. The points representing workload are fitted by the green regression line. It slopes downward, indicating that, with increasing degrees of automation, the routine performance benefits are paralleled by decreasing workload. This highly plausible relationship was predicted in Figure 4.1.

Importantly, the failure response data points (red squares, fitted by the red regression line) also slope downward, showing an opposite trend from routine performance. In any given study, increasing benefits for routine performance with higher automation are accompanied by increasing costs for imperfectly reliable automation, again reflecting the tradeoff predicted earlier. Finally, there were too few data on the effects of increasing degree of automation on situation awareness to observe any well-defined trend, thus not included in Figure 4.4.

A statistical analysis of the above data shows that 12 of the 14 routine performance slopes were positive ($p<.05$; Binomial test). This suggests that investigators were generally successful in their implementation of higher degrees of automation in that more automation increasingly offloaded the operator and consequently, performance of the joint human-machine system was improved. Second, a significant negative correlation between routine performance slopes and workload slopes was observed ($r= -0.72$; $p<.025$), indicating that improved performance with more automation was coupled with reduced workload (see Figure 4.3). Third, while failure performance slopes downward, its correlation with routine performance is essentially zero ($r=0.02$). Although there were few data points for situation awareness, the correlation between situation awareness and both the routine performance slope ($r=0.53$) and the failure performance slope ($r=0.33$) were both positive, as expected.

To examine the effects of LOA more closely, a second analysis was conducted in which the baseline data were removed and only data from the ‘automated’ conditions (Figure 4.2) were analyzed. This analysis results in significantly weaker effects. Now some of the effects were far less pronounced: (1) Only four (of 12) studies showed positive slopes for routine performance; in another four studies, they were less than 0; and the remaining four had slopes around 0. (2) The correlation between routine performance slopes and workload slopes dropped markedly to a non-significant (– 0.23) value. (3) The correlation between routine performance and failure performance slopes was now positive (+0.47), albeit non-significant ($p>0.10$). (4) Despite few data points for situation awareness, the correlation of this measure with all three other measures was now positive, as expected, and quite strong for routine ($r= +0.99$) and failure ($r=+0.96$) performance.

Discussion

Analysis of the results provided two somewhat different pictures, depending on how the analysis proceeded. When the baseline condition was included, such that "more automation" included the introduction of automation, as well as increasing its degree, better routine performance was associated with both lower workload and higher SA. However, failure performance was not associated with routine performance (as mediated by degree of automation), partially at odds with the "conventional wisdom". And automation manipulations that increased SA were associated with increasing performance of both the routine system and the failure response.

When the baseline was removed, increasing degree of automation no longer improved routine performance, and the data now revealed a positive association between routine and failure performance, hence at odds with the conventional wisdom. Critically however, in this second analysis, the data reveal an even stronger relationship between SA and both routine and failure performance, again reinforcing the critical importance of the SA construct.

The predictions based on the "lumberjack analogy" are not well supported by the findings from the second analysis which excluded that baseline data. There are four possible reasons why the expected performance effects of increasing degrees of automation were not observed. First, in seven of the studies, a higher degree of automation did not improve routine performance. This suggests that the independent variable may not have been manipulated successfully. Second, higher degrees of automation, while increasing routine performance in some cases, may actually increase, rather than decrease workload, as it imposes greater automation management and monitoring demands.

Third, with regard to the predicted tradeoff between routine and failure performance, it is possible that additional factors, such as effective automation interface designs (to support situation awareness) and proper operator training (to induce realistic expectations of possible failures; Bahner, Huper & Manzey , 2008), can mitigate, eliminate and even reverse the relationship between both dependent variables (i.e., creating a mild “trade-on” as suggested by the positive correlation of 0.30 reported above). It is difficult to draw firm conclusions, however, given the absence of studies that examined situation awareness in conjunction with failure performance.

Fourth, we note that only three of the studies that contributed to the routine-failure tradeoff analysis employed the highest level of late stage automation. The fact that the other studies did not include this level may well have affected the tradeoff. This is suggested by the possible shape of the tradeoff curve shown in Figure 4.1 which depicts a cost for failure performance only at the highest degrees of automation.

The main conclusion that emerges from the above analysis is that the “conventional wisdom” concerning performance effects of increasing degrees of automation appears to be an over-simplification. A host of contextual factors and moderator variables, such as display design and training, need to be taken into account to be able to predict these effects more accurately.

Chapter 5

Design and Evaluation of Dynamic Function Allocation Schemes and Associated Interfaces

Introduction

The experiment described in Chapter 3 and the meta-analysis described in the previous chapter served to highlight the negative impact of higher level of automation on human performance in automation failure scenarios due to automation complacency (e.g. Crocoll & Coury, 1990; Sarter & Schroeder, 2001). One solution to this side effect of automation is to design a system that is flexible and responsive to user needs, environmental demands, and context.

Context-sensitive automation traditionally takes the form of either adaptive (i.e., system-controlled) or adaptable (i.e., user-controlled) changes to the status and behavior of the system (Opperman, 1994). Adaptable automation gives users (a sense of) control and can increase awareness of the system state (i.e., the level at which the automation operates and any changes to those levels). As a result, this form of automation tends to be preferred by operators. However, the drawback of adaptable automation is that operators may experience higher workload due to the additional automation management task, and they may fail to choose the appropriate level of automation for a given context. Adaptive automation, on the other hand, reduces the burden on the operator but does so at the risk of confusion over assigned responsibilities. Past research has shown that participants often rate this form of context-sensitive automation as less desirable than adaptable automation. The effects of adaptive automation are well documented (Cosenzo, et al., 2006, Kaber & Endsley, 2003; Calhoun Ward, & Ruff, 2011), so are those of adaptable automation, although to a lesser extent (Miller, Funk, Goldman, Meisner, & Wu, 2005; Parasuraman, Galster, Squire, Furukawa, & Jiller, 2005).

In this chapter, we describe a hybrid system, in which the system and the operator collaborate on selecting and activating appropriate automation levels. The default mode is adaptive but operators can engage in automation management and override system selections when they feel the need and have the time to do so. We expect that the hybrid system results in the best overall system performance as it combines the benefits of both adaptive and adaptable approaches. In particular, we anticipate a fairly high level of perceived control (higher than with adaptive and slightly lower than with adaptable automation). Operator workload is also likely to fall between the one experienced with adaptive and adaptable automation. User awareness of automation state is expected to be as high as with adaptable automation.

In this final experiment, these predictions are being tested as we compare the effectiveness of the hybrid system to fixed function allocations and to an adaptive and an adaptable function allocation scheme.

Method

Participants

Twelve University of Michigan students (4 female, 8 male; age 20-28) who participated in the earlier studies that were described in Chapters 2 or 3 of this document returned for this experiment. Their prior experience with robotic arm control task ranged between 6-9 hours, which included the training session(s) and the experiment session from the previous study. These participants were recruited because their relatively extensive experience (compared to first-time participants) was more comparable to that of experts, and thus made the findings more generalizable to real-world operations. Also, involving these participants somewhat reduced the training requirements in this final study. All participants reported normal or corrected-to-normal vision (only contacts are acceptable due to limitations of the eye tracker). Participants were compensated at a rate of $15/hour.

Experimental Design

The independent variable in this study was the type of function allocation scheme (fixed, adaptive, adaptable, or hybrid). A within-subject factorial design was used. The first condition was fixed automation, in which all stages of automation (stage 1&2 represented by hazard avoidance, and stage 4 represented by trajectory control) were set at a medium level. The medium level was chosen to avoid the high costs of the highest level based on the results from Chapter 3. The second condition was adaptive automation, in which the levels of automation were adjusted by the system based on observed operator performance (Calhoun, Ward, & Ruff, 2011). The performance measure used for the hazard avoidance task was 'proximity to hazard'; the performance measure for the trajectory control task was the 'deviation of the EE position from the ideal path". The third condition was adaptable automation, in which the operator was in charge of setting the levels of automation. Finally, the fourth condition was a hybrid automation scheme that combined adaptive and adaptable automation. The system constantly evaluated the operator's performance. If, at the manual or medium level, performance deteriorated, the system moved to the next higher level (medium or high level of automation, respectively). However, the participants were able to choose their own preferred level of automation and override the one chosen by the system.

In all conditions, the default setting of the system was medium for both hazard avoidance automation and trajectory control automation. In the dynamic automation schemes, i.e. adaptive, adaptable, and hybrid automation, hazard avoidance and trajectory control automation levels could change independently of one another. Table 5.1 shows detailed descriptions of all three levels of each automation function, as well as the behavior of the three automation schemes.

In order to examine human interaction with different automation schemes, the experimental scenarios and the mechanism of dynamic function allocation should be designed in such a way that different automation schemes indeed worked differently in the sense that:

(1) The higher level of automation will be triggered in the adaptive and hybrid schemes (to ensure that they are different from the 'fixed' automation scheme at the medium level).
(2) The operator has a motivation/incentive to override the level of automation in the hybrid scheme (to ensure that the hybrid scheme is different from the adaptive scheme). (Note that it is unlikely that the operator will voluntarily use the manual mode without any hazard avoidance or trajectory guidance).

Both aspects were addressed, to some extent, in each of the two types of automation. For hazard avoidance automation:

(1) All tasks involved a geometry that brought the arm/EE close to the table. Some of the tasks also involved situations where the arm configuration was close to (but not in) a warning zone.
(2) A high level of hazard avoidance leads to a reduced likelihood of violating flight rules or colliding with structures. The medium level does not guarantee that violations of flight rules or collisions are avoided but it helps avoid abrupt pauses during the operation and thus saves time. In other words, each level has its benefits and disadvantages, and therefore the operator may be motivated to use both.

For trajectory control automation:

(1) Flying diagonally across the table involves multi-axis hand control and is therefore difficult for novice operators to perform. At the medium level of trajectory control automation, the trajectory error is unlikely to be zero at all times. This increases the likelihood that a higher level of automation will be triggered.
(2) The difficulty of flying diagonally is the motivation for the operator to override the automation and select a higher level. However, during training, participants

will experience failures of the trajectory automation that lead to hazards (like on trial 2 in Experiment 1). This may motivate them to stay at the medium level as much as possible where it is easier/faster to respond and recover when hazard alerts are encountered. In addition, participants will be told at the beginning of the experiment that failures may occur even though none will be included in the actual experimental scenarios.

Table 5.1: Description of automation functions

		Hazard avoidance		***Trajectory guidance or control***
		Stage 1: Information acquisition	***Stage 2: Information integration (analysis)***	***Stage 4: Action implementation***
Subtask		*Monitoring the configuration of the arm while moving the arm*	*Monitoring clearances (i.e. analyzing the relevant location of the arm with respect to obstacles/structures) while moving the arm*	*Operating (moving) the arm*
Functions at different levels	High	Automation provides a graphic indication, as well as stops the arm to prevent going further into the problematic configuration when it is being approached (e.g., within 5 degrees).	Automation provides a graphic indication, as well as stops the arm to prevent going further into a collision (or violation of flight rule) when the arm is within 0.8m of structure.	Auto control: Automation flies end effector automatically along the ideal trajectory; a visible path is presented in the camera views.
	Medium	Automation provides a color-coded graphic indication on the visual cues (monitors) of the joint in which a singularity is expected to occur when it is being approached (e.g., within 5 degrees).	Automation provides a graphic indication where 'violation of the flight rule' is expected to occur (e.g. within 0.8m of structure, i.e. within 0.2m of flight-rule-distance).	Auto guidance: Automation provides advice for hand controls to fly the ideal trajectory; a visible path is presented in the camera views.
	Manual	Human operator monitors for potential problematic configurations.	Human operator monitors for potential clearance issues.	Human operator manually operates the arm. The ideal path is presented in the digital map (to match the visible path in the camera views for medium- and high-level automation).
Automation Schemes	Fixed	Medium level for both stages (both collision avoidance and arm configuration alert)		Medium level

	Adaptive	The baseline is the medium level of automation; the high level of automation is triggered by poor performance. Automation switches back to the lower level if performance improves and stays at the lower level for an extended time. Automation will not switch to 'manual' under any circumstances. Take the collision avoidance as example, when the arm is 0.8m from a structure, there will be a warning. If the operator keeps going closer to the structure, and the clearance falls below 0.6 m (e.g. 0.5m), the higher level automation will be triggered, i.e. the automation will stop the arm. If the operator backs out and the clearance increases to more than 0.6 (e.g. 0.7m) and stays at this or a larger distance for 5 seconds, the automation will switches to lower level.	The baseline is the medium level. The high level is triggered by poor performance (big trajectory error, e.g. 1m). Automation switches back to medium level if the trajectory error decreases to 0.6m and stays at that level for 5 seconds. Automation will not switch to 'manual' at any circumstances.
	Adaptable	The initial setting is the medium level of automation (to be comparable with other schemes). Human operator is responsible for initiating any changes to the level of automation.	The initial setting is the medium level of automation. Human operator is responsible for initiating any changes to the level of automation.
	Hybrid	The default is adaptive: the initial setting is the medium level; the high level is triggered by poor performance. The operator can override the level of automation (adaptable) to go to any of the three levels. If at the manual or medium level, the performance deteriorates, the system then goes back to one level higher (this will not become in effect until 10 seconds after a change to the automation level is initiated by the operator). The override will be disabled within 10 seconds of a change initiated by the system (the buttons on the system state GUI will look different to indicate that they are not available/no override is allowed) .The buttons will also change color to indicate that a change was made by the system ("blue" as used across other BORIS GUIs) or by the operator (a "warmer" color, orange)/	The default is adaptive: the initial setting is the medium level; the high level is triggered by poor performance. The operator can override the level of automation (adaptable) to move to any of the three levels. At the manual or medium level, if the performance deteriorates, the system then goes back to one level higher (this will not take effect until 10 seconds after a change of the automation level initiated by the operator). The override will be disabled within 10 seconds of a change initiated by the system.

The dependent measures in this study included performance outcome measures, process measures, mental workload assessments, and subjective preferences for automation schemes. The performance outcome measures were:

(1) Task completion time for the entire task.

(2) Accuracy. Mean deviation from the desired path, and from the target final position and orientation of the end effector.

(3) Hazard occurrence and recovery from hazard. The number of times a hazard was encountered, and the required time for and accuracy of recovery from the hazard.

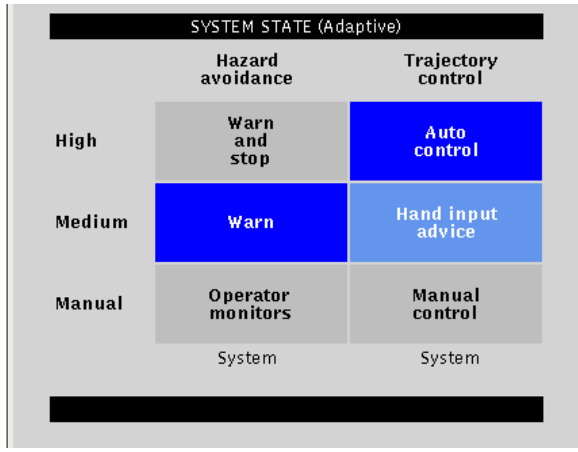
The process measures included the management of automation levels, operators' awareness of automation configuration and system-induced adjustments, and the operators' scan pattern and effectiveness. Management of automation levels refers to when and how the operator initiated changes to the level of automation in the adaptable and hybrid schemes, and how the system changed the level of automation in the adaptive and hybrid schemes. The awareness of the system state was determined based on (1) the ability of the participant to recall the system states throughout a scenario at the end of that very scenario; (2) button presses to indicate the detection of a system-induced change to the automation level in the adaptive scheme. The scan pattern and effectiveness were examined using eye tracking data, which was collected using an ASL Eye-Trac 6D eye tracker at a sample rate of 60 Hz. Fixations and dwells in areas of interest (AOI) were analyzed. The AOIs included the camera views and the window view, the system state GUI, the joint angle display, and the camera selection area.

Other dependent measures included mental workload (on a 0-10 scale), which was measured after completion of each task, reported strategies for the arm operation and management of automation levels, as well as perceived advantages and disadvantages of the automation schemes and interfaces, as gathered through a debriefing questionnaire.

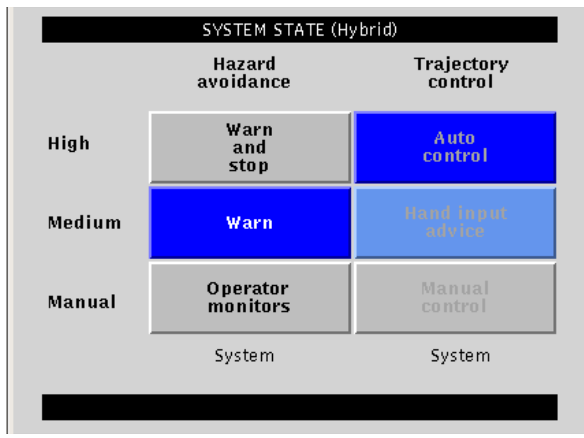
Simulation

As in the earlier experiments, we used the BORIS simulation which was, in this case, augmented by the addition of a new GUI to support at-a-glance monitoring and control of the automation configuration. The layout of the GUI was consistent across all automation schemes: three levels for each stage of automation were presented using a 3 (rows: high, medium, low) x 2 (columns: hazard avoidance, trajectory control) matrix. Each cell was labeled with a short description of the function that the specific stage and level of automation would perform. In the adaptive automation scheme (see Figure 5.1(a)), the cells showing the active level for each stage of automation were highlighted in blue. This color was chosen to be consistent with other GUIs in the system. When the system initiated a change to a level of automation, the cell including the new level turned blue and flashed for 10 seconds. The immediate past level of the automation was shown in light blue for 15 seconds. The highlighting of current and past levels supported at-a-glance monitoring and awareness of current and past levels of automation for both tasks.

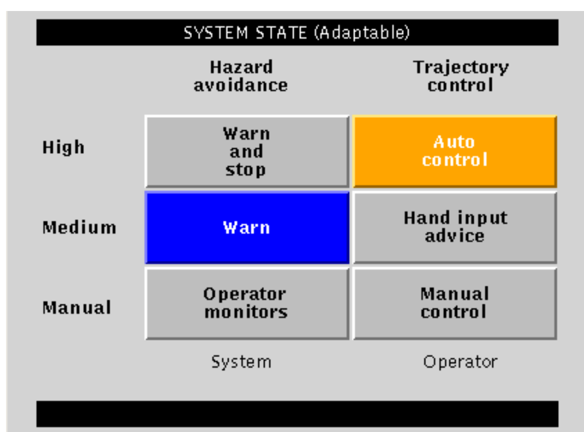
In the adaptable and hybrid schemes, the GUI was not only a display of the system state, but also the control panel for the operator to set/change the level of automation. Each cell turned into a button that the operator could press to activate the corresponding automation setting. If the operator (rather than the system) was the initiator of a level change, the cell would be highlighted in orange (as opposed to blue; see Figure 5.1(b)). Redundantly, at the bottom of each of the two columns, there was a text-based indication showing which agent (operator or system) initiated the last change to the level of automation for this stage. In the hybrid system, an adjustment period of 5 seconds was imposed after a change was initiated. During that time, the level of automation for that specific task could not be changed by either the system or the operator. During this period, the buttons of that specific stage were disabled which was indicated to the operator by changing their appearance to a faded look (see the three buttons for trajectory control automation in Figure 5.1 (c)) which indicate "non-executable"). The enabled and disabled appearances of the buttons were designed to provide affordance (Norman, 2002), thus ensure safe and efficient shift.



(a) Adaptive



(b) Adaptable



(c) Hybrid

Figure 5.1: System State GUI for (a) adaptive, (b) adaptable, and (c) hybrid automation schemes.

Tasks

As in the earlier studies, the participants played the role of a payload specialist and were responsible for completing fly-to tasks, each of which involved planning a trajectory, choosing and adjusting proper camera views, and operating the arm safely with two hand controllers to reach a target EE position and orientation.

The goal of the task was to move the arm to the target position as safely and efficiently as possible. Safe operation was defined as keeping sufficient clearance between the arm and any surrounding object/structure, and staying away from problematic arm configurations such as singularities and reach limits. Efficiency referred to flying the arm at the coarse rate as much as possible and along the shortest trajectory (for example, moving across the table diagonally). Also, participants were asked to make sure that they reach the EE target position within given tolerance limits (accuracy).

The participants were instructed that safety, accuracy, and efficiency were equally important to encourage them to use, as much as possible, the medium level of automation. For example, one disadvantage of high-level of automation was that overall task completion took longer time if the EE went too close to a hazard and was stopped by the hazard avoidance automation. The 3 participants with the best performance, in terms of accuracy (e.g. reach the target point within tolerance), efficiency (e.g. take the shortest trajectory and use the coarse rate as much as possible), and safety (e.g. avoid problematic arm configurations and collisions), received a $20 bonus in addition to the $15/hr compensation.

In addition to the fly-to-tasks, participants were instructed to indicate the detection of a system-initiated change to an automation level by pushing a button on the rotational hand controller. They were also asked to verbalize the type of change as soon as possible.

Procedure

The participants completed one training session of 3-4 hours on the first day and then one experiment session of 3-4 hours on the following day. In the training session, the participants completed the following 7 scenarios:

- two scenarios without any automated aids (scenarios 1 and 2);
- one scenario with medium-level hazard avoidance automation and medium-level trajectory control automation (trajectory guidance; scenario 3);
- one scenario with high-level hazard avoidance automation and high-level trajectory control automation (scenario 4);
- one scenario with adaptable, adaptive, and hybrid automation schemes, respectively (scenarios 5, 6, and 7).

Scenarios 4, 5 and 7 all involved trajectory automation failures. In scenario 4, the trajectory control automation flew the arm along the shortest path (connecting the start point, two way points right above the start and end point, and the end point) which resulted in the elbow getting closer than 0.6m to the FORWARD wall. As a result, the high-level hazard avoidance presented a warning and simultaneously stopped the arm.

In scenario 5, the high-level trajectory control automation (if chosen by the operator) left the correct path and ran into an elbow pitch singularity. If the participant did not choose the high level for trajectory control automation, the experimenter switched the level of trajectory automation to high and demonstrated the failure to the participant after the participant completed the scenario.

In scenario 7, the high level trajectory automation led the arm into a self-collision warning at the beginning of the scenario. The participant had to switch to the medium level of trajectory automation and fly the arm to temporarily reduce the wrist yaw angle and correct the same joint angle later in the scenario (either with medium-level trajectory automation or high-level automation).

The purpose of introducing these failure scenarios was to make participants aware of the risk of automation failures, especially when operating at the highest level of automation, and thus to avoid that they stay at that level at all times during the actual experiment. This precaution was taken based on pilot tests where some participants showed that very tendency. The experimenter also verbally encouraged the participants to use lower-level automation as much as they could by emphasizing the high costs of failures and hazards in the real world.

All participants were able to follow the correct procedure to operate the arm to complete the fly-to tasks, and to manually fly the arm to the target point within tolerance (1 meter for positions, 15 degrees for orientation) at the end of the second training scenario. Also, all participants had a good understanding of the behavior of the three automation schemes at the end of training scenario (7).

In the experiment session, each participant completed 2 blocks of trials. There were 4 trials in each block, which consisted of:

- Normal scenario with fixed automation scheme.
- Normal scenario with adaptive automation scheme.
- Normal scenario with adaptable automation scheme.
- Normal scenario with hybrid automation scheme.

Eight geometries of the trajectory for fly-to tasks were designed for this experiment. All geometries consisted of three segments: up, across (over) the table, down. The order of the geometries used in trials 1-8 was fixed. The order in which the automation schemes were encountered by the participant in these trials was counterbalanced.

Results

Performance Measures

Repeated measures linear models (using the General Linear Model formulation in SPSS 20) were used to analyze the main effects of the automation scheme. Two-tailed Fisher's LSD post-hoc tests were then conducted on significant results to determine which conditions differed significantly from one another.

Completion time. Figure 5.2 shows completion times for fly-to tasks for the four conditions (average of two trials for each participant under each condition). The effect of automation scheme on completion time was not significant ($F(3)$=2.529, p=0.074).

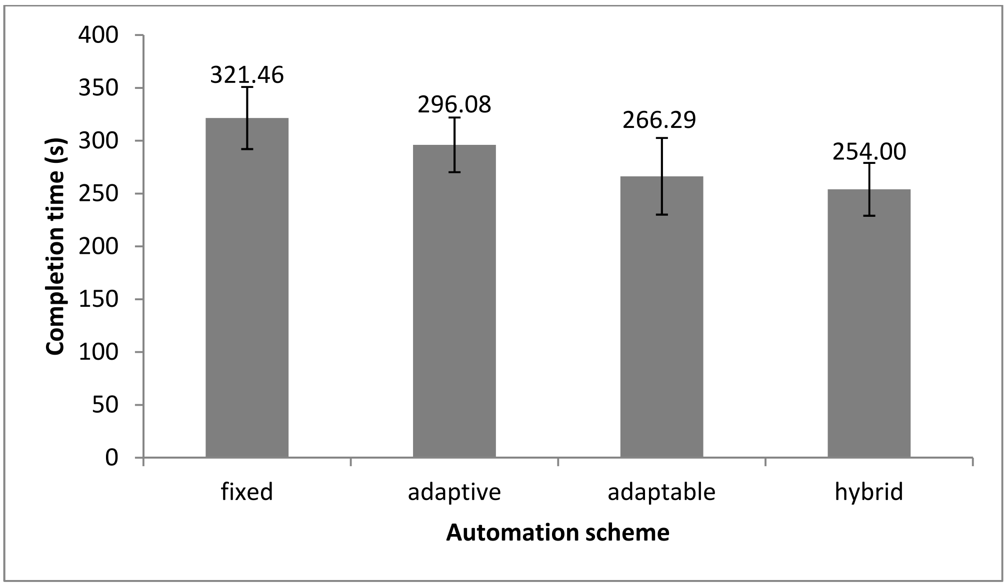


Figure 5.2: Completion time for fly-to tasks as a function of automation scheme. Error bars represent standard error

Trajectory deviation. Trajectory deviation, i.e. the average distance between EE and the optimal path throughout a trial, differed significantly as a function of automation scheme (F(3, 9) = 5.058, p=0.025, see Figure 5.3). Post-hoc analyses showed that trajectory deviation in the fixed automation (M=23.03, SD=19.24) was significantly larger than in all three dynamic automation schemes: the adaptive scheme (M=12.44, SD=4.76, p=0.05), the adaptable (M=9.42, SD=7.69, p=0.03), and hybrid condition (M=8.48, SD=3.99, p=0.02). In addition, trajectory deviation in the hybrid scheme was significantly smaller than that in the adaptive condition (p=0.007).

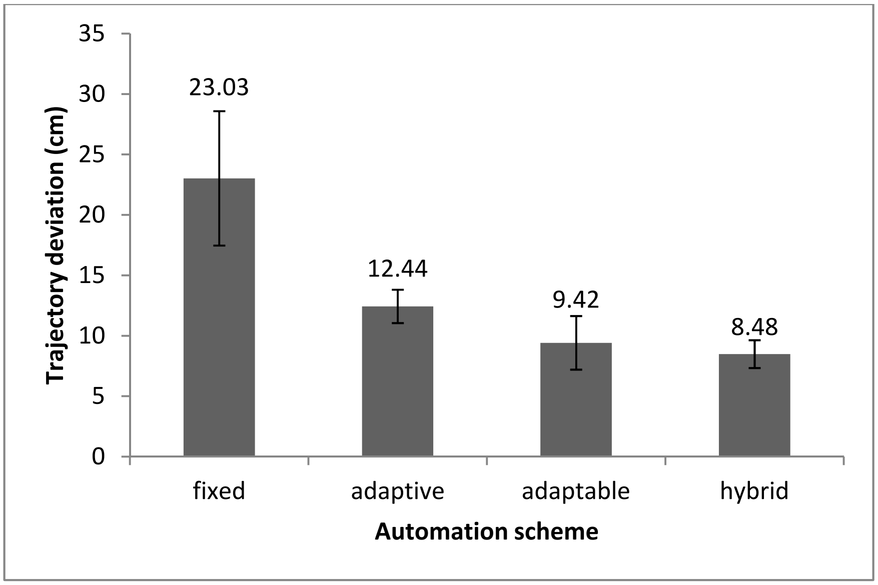


Figure 5.3: Trajectory deviation as a function of automation scheme. Error bars represents standard error

Participants' strategies for choosing automation levels in the adaptable and hybrid schemes

Trajectory control automation. In the adaptable and hybrid conditions, participants used one of two basic strategies for setting the automation level for trajectory control. One group preferred to stay at the same level throughout an entire trial while the other group switched between levels. We will refer to the latter group as 'mixed'. Within the former group, some of the participants stayed at the high level (auto control) at all times while others preferred the medium level (auto guidance). None of the participants ever switched to the low level (manual control) of trajectory control automation.

Table 5.2 shows for trials that started at either the medium or high level of automation (12 each in the adaptable condition, and 14 and 10, respectively, in the hybrid condition), how many times participants stayed at the medium or the high level of automation throughout the trial and how many adopted a 'mixed' approach. Chi-square tests showed no significant difference between the two conditions.

Table 5.2: Participants' choices of (changes to) the level of automation for trajectory control (24 trials for each condition)

	Initial setting			Throughout scenario	
	Adaptable	Hybrid		Adaptable	Hybrid
Medium	12	14	Medium throughout entire trial	8	10
			Mixed of medium and high	7 (3 high->medium) (4 medium->high)	6 (3 high-> medium) (3 medium->high)
High	12	10	High throughout entire trial	9	8

Among the 12 participants, four used the high level of trajectory control automation in all four trials in both the adaptable and hybrid conditions; another four participants tended to use a combination of both the medium and high levels; and the remaining four used medium throughout the entire scenario for the four trials. Based on an analysis of the participants' comments during the scenarios and their responses to the debriefing questionnaire, it appears that those who preferred and used auto control throughout the scenario perceived the high level as more efficient and accurate. For example, one participant said, "Automation is most efficient, and does better than I do. I can still watch". This group of participants believed, incorrectly, that, even if a failure occurred (e.g. high level automation flew the arm towards a hazard and the arm was stopped by the high-level hazard avoidance), the overall completion time would still be shorter than if the trajectory was controlled by the human operator. In addition, they did not consider automation failures to be critical events because they relied on the high level of hazard avoidance to stop the arm from colliding with a structure or obstacle. For example, one participant said: "I can take over if it stops", while another said: "if the automation can't finish, I will take over". For these participants, the hybrid and the adaptable schemes ended up working nearly the same way. In both cases, the scenarios were completed flawlessly with auto trajectory control since no failures were introduced during the actual experiment.

In contrast, those who preferred the medium level of automation did so because of concerns over failures at the high level and because they considered their own ability to control the arm to be “good enough”. They also explained that they enjoyed being in control or wanted to be the one who “makes decisions”. These participants were indeed quite proficient at controlling the arm and performed well with the help of auto guidance. As a result, only two of the 16 adaptive and hybrid trials performed by the four participants in this category saw any system-initiated changes to the automation level.

Three participants preferred to use auto control only for the difficult portion of the scenario. For example, two of the three participants considered multi-axis hand control to be difficult and therefore used auto control only during the second segment of the trajectory when the arm needed to move diagonally over the table. The third participant considered EE rotations to be most challenging and therefore used auto control only for the first segment which required moving the arm up and rotating the EE simultaneously. Using different levels of automation for different phases of a scenario accounts for most observed changes to the level of automation in the middle of a scenario. Other reasons for mid-scenario changes included: (1) switching to the high level during the third segment of a scenario because the EE got very close to the table and AFT wall; (2) switching to the high level to allow the auto control to do the final adjustments at the end of the task; and (3) switching to the medium level at the end of the task because the EE was not perfectly aligned with the target point.

Participants employed different strategies also in response to encountering a warning (either while using auto control or auto guidance). Two participants (each encountered a warning in one trial) completed the rest of the trial using auto guidance (manually controlling the arm) because they were worried that the auto control was not capable of performing the task reliably. In contrast, another participant chose to switch to the high level after encountering multiple warnings because he failed to diagnose the cause to the warning and was not able to get away from the warning.

Hazard avoidance automation. For hazard avoidance, the majority of participants chose to stay at either the medium or the high level of automation throughout a trial (see Table 5.3). The 'mixed' approach was used in only one trial. A strong correlation was found between the level of trajectory automation and the level of hazard avoidance initially set by the participants (r= 0.873, p<0.001), as well as between the level (medium, mix, high) of the two automation functions through the entire trial (r=0.712, p<0.001). The high level was chosen most often when the target position of the EE was close to table.

Table 5.3: Participants' choices of (changes to) the level of automation for hazard avoidance (24 trials for each condition)

	Initial setting			Throughout scenario	
	Adaptable	Hybrid		Adaptable	Hybrid
Medium	12	14	Medium throughout entire trial	12	14
			Mixed of medium and high	1	--
High	12	10	High throughout entire trial	11	10

System initiated changes and operator responses

Table 5.4 shows a summary of system-initiated changes in the adaptive and hybrid conditions, as well as participants' responses to these changes. In all adaptive trials that involved a system-initiated change from the medium to the high level of automation, such a change occurred only once since, for the remainder of the trial, the operator's performance remained above threshold for the rest of the trial. In contrast, in 3 out of the 5 trials in the hybrid condition, a system-induced change from the medium to the high level of automation occurred twice due to repeated poor operator performance. In one instance, the participant mistakenly referred to the wrist roll joint instead of the end point of the EE as the origin of FOR (frame of resolution). As a result, the participant deviated from the path and triggered a change from the medium to the high level of automation. Since she did not realize her mistake, she was convinced that the auto control system made a mistake. Therefore, she overrode the system by changing the trajectory automation back to the medium level. This repeated once. In the other two cases that involved two changes from the medium to the high level, the participants simply deviated from the path during normal operation.

Noticing time (reaction time to a system-initiated change) was defined as the time between when the system initiated the change and the time when the participant pressed the button on the rotational hand controller. A repeated-measure ANOVA showed that reaction times to switches from high to medium (M=3.34, SD=1.17) were significantly longer than reaction times to switches from medium to high (M=2.35, SD=1.02), $F(1, 10)=15.840$, $p=0.003$.

Table 5.4: Summary of system-initiated changes

	Adaptive	Hybrid
Trials in which the medium level was used for the entire or for parts of the scenario	24	16
Trials where the system initiated changes to the automation level	8	5
Switches from medium to high by the system	8	8
Switches from high back to medium by system	8	6
Total number of switches by system	16	14
Switches where participant pressed button to acknowledge that they noticed the change	15	11
Reaction time (medium->high) (s)	M=3.28	M=3.42
Exclude changes where the participant only responded to one of the switches	SD=1.18	SD=1.30
Reaction time (high->medium) (s)	M=2.49 SD=1.32	M=2.15 SD=0.37

Note: The participants failed to acknowledge some level changes, but they did not necessarily fail to notice them. In one of the cases, for example, the participant was adjusting the camera views after the auto control took over, thus when it switched to the medium level, her hand was on the mouse instead of the joystick and thus did not press the button. However, in her response to the questionnaire after that trial, she recalled the change from high to medium correctly.

The use of high-level automation in the adaptive and adaptable schemes

The observed frequency of using the high level of trajectory control automation was different for the adaptive and the adaptable schemes. In the adaptable condition, 16 (7 mixed and 9 high) of the 24 trials were completed using the high level of trajectory control for part of or for the entire trial; in contrast, in the adaptive condition, the high level trajectory automation was employed in only 8 of the 24 trials. A chi-square analysis showed a significant difference between these proportions (χ^2=5.33, p=0.021). Table 5.5 lists the number of trials in which medium, high or both levels were used under the two conditions.

Table 5.5: Participants' or the system's choices of (changes to) the level of automation for trajectory control in adaptive and adaptable conditions (24 trials for each condition)

	Adaptive	**Adaptable**
Medium throughout entire trial	16	8
Mixed of medium and high	8	7
High throughout entire trial	--	9

Percentage of time spent at high level of automation

The percentage of time spent at the high level of trajectory automation was very high and significantly longer in the adaptable (p=0.011) and hybrid (p=0.017) conditions than with the adaptive scheme ($F(2)$= 7.528, p=0.004.; see Figure 5.4).

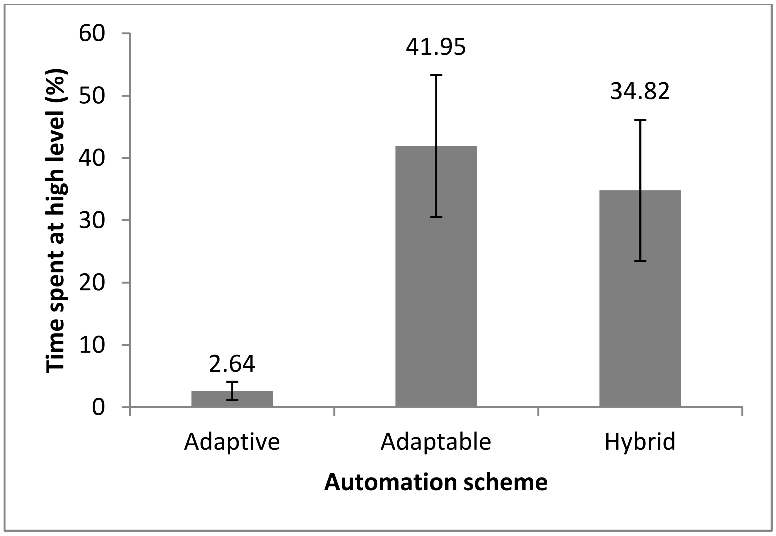


Figure 5.4: Percentage of time using high-level trajectory control automation per scenario

System awareness (measured by the recall task)

Participants' recall of changes to automation levels during each trial was compared with the actual changes. The recall score was set to be 1 point if it was completely correct, 0.5 points if partially correct (e.g. if two changes were initiated, but recalled one of them), and 0 points if completely incorrect . The average scores for the four automation schemes were all above 90% (100% for fixed and adaptive conditions, 91.7% for adaptable and hybrid schemes) and did not differ significantly from each other.

Subjective rating on ability of maintaining awareness of the system state

The participants were asked to rate their ability to track the automation levels throughout each trial. The average ratings for all four conditions were relatively high (above 7.8 on a 0-10 scale, see Figure 5.5), and did not differ significantly from each other.

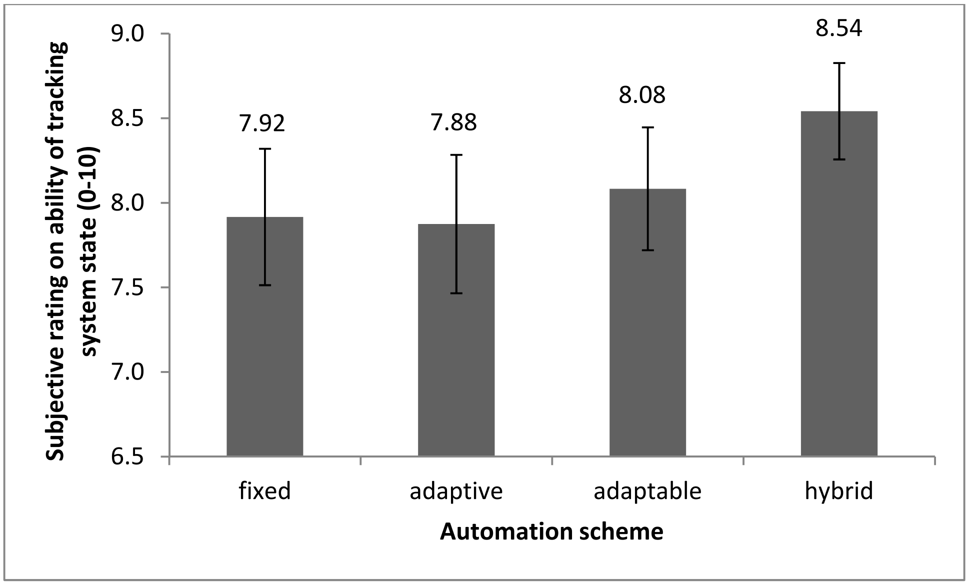


Figure 5.5: Subjective rating on ability of tracking the system state. Error bars represent standard errors

Subjective workload

Mental workload varied significantly as a function of automation scheme, F(3)=5.938, p=0.002 (see Figure 5.6). Pairwise comparisons showed that the fixed scheme resulted in higher perceived workload (M=4.42, SD=0.48) than both the adaptive scheme (M=3.83, SD=0.33, p=0.046), the adaptable scheme (M=3.17, SD=0.28, p=0.021), and the hybrid scheme (M=3.29, SD=0.43, p=0.006). Also, mental workload in the adaptive case was higher than in the adaptable (p=0.039) and hybrid (p=0.035) conditions. The difference between the adaptable and hybrid schemes was not significant.

The analysis of physical workload also found an effect of automation scheme (F(3, 9)=5.542, p=0.020). Pairwise comparisons showed significant differences between fixed scheme and adaptable (p=0.027) or hybrid scheme (p=0.003), and a significant difference between adaptive and hybrid schemes (p=0.002).

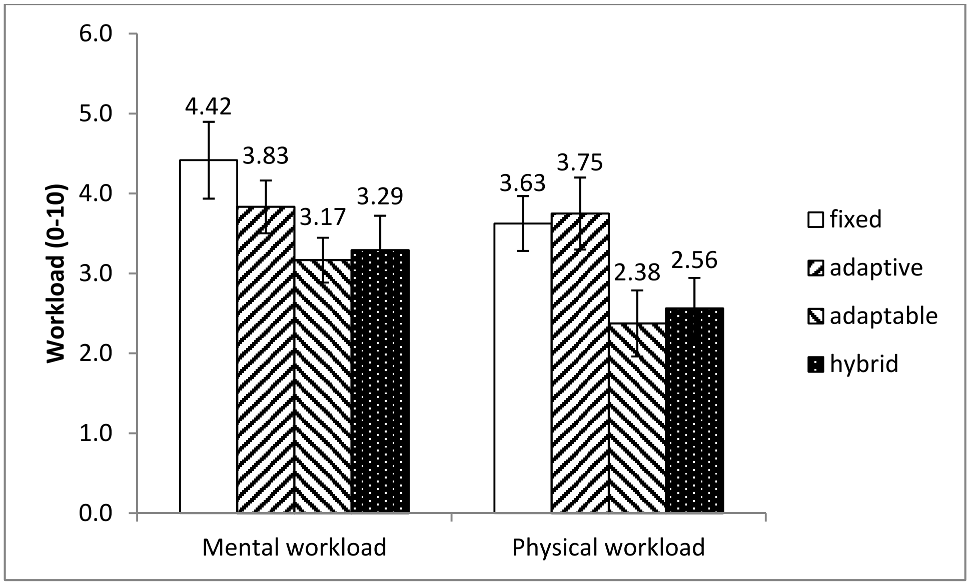


Figure 5.6: Mental and physical workload for four automation schemes. Error bars represent standard errors

Preferences for automation schemes

Seven participants considered adaptable automation to be the most preferable scheme, five chose the hybrid scheme, and none preferred adaptive automation.

The adaptable automation scheme was preferred because it gave the operator full control. Quotes from the participants include: “I could decide what I wanted to do and didn’t have to have the system change on me if I didn’t want to”. “I liked being the one changing the system state.” “I enjoyed having the control over the system. It was easy to manipulate the system into doing what I thought was correct.”

Operators who preferred the hybrid scheme explained that it provided them with a sense of control over system state, in combination with the system monitoring their performance. For example, one participant said: “I prefer knowing that if I went too far off course, the system would be able to provide guidance, but I would be able to override it.” Another participant stated: “I liked the hybrid scheme. It was as if there was someone double checking my actions for me.” Some participants would have preferred the adaptive part of the hybrid system to be a suggestion rather than an automatic switch: “If the state-change notification were more of a suggestion and more prominently displayed, I would also like the hybrid.” Another participant commented on the hybrid scheme with a possible automation failure in mind: “Hybrid might have also been a good scheme as it keeps the arm moving in the projected motion. The only issue with this is that the projected motion might not always be possible/optimal. That is where human interference can come in and change it up”.

Concerning the adaptive scheme, most participants felt that the changes initiated by the system made sense. However, they also mentioned a range of problems with the adaptive automation:

(1) The operator could not tell/predict when a change was going to happen. “I was always surprised”, said one participant. Participants suggested that some feedback be added to indicate the EE’s deviation from the path, or some warning be triggered before a change was initiated.

(2) The threshold for triggering the change was not considered appropriate by all participants. For example, one participant said “sometimes it wasn’t necessary for it to change to auto because I didn’t think I was too far off the path.” Another participant said, “It was a little picky, though”.

(3) The operators may have reasons for deviating from the path that the automation is not aware of. For instance, one participant said he deviated from the path intentionally in order to ensure clearance from the table.

(4) Four participants stated that the system-initiated changes were sometimes confusing, particularly when the change was not what the participant expected or wanted to do. In addition, they were not sure whether the flashing of the corresponding GUI cell indicated that the trajectory automation was going to, or had already taken over. The detection rate for changes initiated by the system was high, though. Only one participant thought that they had missed changes to the system state. He commented: ”It was hard to take my eyes away from the

arm itself when something went wrong. I found myself looking more at the highlighted bar than at what the system was doing in the background".

System State GUI

Eleven participants responded to the question "Did you find the system state GUI helpful? Do you have any suggestions for improvements?" Ten participants indicated that the GUI was very helpful and efficient. They felt that the GUI presented the stages and levels of the automation in an intuitive format. Importantly, participants found it easy to spot any changes to the automation levels, and they considered the salience of the flashing cell in case of system-initiated changed to be appropriate. Only three participants suggested that it should be made more noticeable, especially during high workload.

Discussion

This study compared joint system performance with fixed automation to three dynamic function allocation schemes: adaptive, adaptable, and hybrid (a combination of adaptive and adaptable) automation. Deviations from the optimal trajectory were smaller with all three dynamic schemes, compared to fixed automation. Within the dynamic schemes, performance with the hybrid approach was superior to that in the adaptive condition. The automation scheme also affected perceived mental and physical workload. Here again, the highest perceived cognitive workload was reported for the fixed automation scheme. Mental workload in the hybrid and adaptable conditions was lowest, followed by the adaptive scheme. Similarly, physical workload was lowest also in adaptable and hybrid schemes.

The performance benefits observed in the adaptable and hybrid schemes can be explained, in part, by participants' rather excessive use of higher level trajectory control automation, rather than by the automation scheme per se. In more than one third of all trials (9 out of 24 trials) with adaptable automation, participants switched to the high level right at the beginning of the scenario and stayed at that level throughout the trial. In another 7 trials, the high level of automation was invoked part time. Thus, altogether, two thirds of the trials in this condition were completed at the high level of automation for at least some time. Similar trends were found for the hybrid scheme. In contrast, in adaptive automation, the high level of trajectory control was activated in only one third of all the trials (8 out of 24 trials) - 50% less than in adaptable automation. As we saw in Chapter 3, a higher level of trajectory control automation (as used extensively in the adaptable and hybrid conditions in this experiment) led to enhanced performance and lower mental workload during routine scenarios. This finding is consistent with findings from a previous study on adaptable and adaptive automation for supervisory control of autonomous vehicles (Kidwell,Calhoun, Ruff, & Parasuraman, 2012). In that study, an automated aid was designed to support an image analysis task at three different levels of autonomy. The default level was 'low' for both adaptable and adaptive automation. In the adaptive automation scheme, changes to a higher level of automation were initiated based on operator performance. The authors found that participants spent significantly more time in the low level of automation in the adaptive condition (21.1%) than with adaptable automation (5.7%). The effects of automation scheme on performance were mixed, however. Performance on the image analysis task was superior with adaptive automation, compared to adaptable automation; whereas the automation scheme has an opposite effect on the performance on a change detection task that was not directly supported by the automation.

To mitigate the confounding factor of level of automation in the current experiment, participants were grouped according to their automation selection strategy: medium, mixed, and high. A 3 (operator style: medium, mixed, and high) x 4 (four automation schemes repeated measures ANOVA showed that neither the main effect of operator style, nor the interaction between operator style and automation scheme, was significant. Note, however, that this may be due to the rather small sample size in this experiment.

System awareness (measured by both performance of pressing the acknowledgement button and the recall task) and the perceived ability to track (changes in) system state were high across all conditions. This could be explained by a ceiling effect. However, there were two special cases in which the participants had a high fatigue level or workload level and, as a result, lost awareness of the task or the system state

Another concern with adaptable automation, i.e. increased workload, was also not observed in this experiment. Both the high level of system awareness and the absence of an increase in perceived workload may be attributable to the low level of task difficulty in this study which led to near- perfect performance and a low frequency of switches between levels of automation. For example, as described in Chapters 2 and 3, the medium level of hazard avoidance was sufficient for preventing hazard encounters and thus the high level of hazard avoidance was rarely triggered in the adaptive or the hybrid schemes. Another reason why the expected disadvantages of adaptable automation were not observed could be the design of the System State GUI. The GUI was considered by participants to be very effective both for supporting at-a-glance monitoring of the system state, and for safely and efficiently shifting between system states. The success of this design verified and specifies the design guidelines proposed by previous research (Kaber, Riley, Tan & Endsley (2001). At-a-glance monitoring can be supported by highlighting the active or selected levels/stages of automation. External attentional guidance can be provided using salient indicators (flashing) when shift from one state to another (in adaptive and hybrid scheme). Interface style consistency can be realized by using the same layout of interface for all automation schemes. Direct manipulating interface (DMI) can be used to support safe and efficient shifts. On the DMI, all available levels of stages of automation should be visible, when a shift is not available, e.g. during the transient phase in the hybrid scheme, the buttons on the DMI should be disabled and appearance of the button should show the status. Warnings of approaching performance limit can be provided by visual notifications that are integrated on the direct visual view of task (in this experiment, the camera views of the robotic arm).

While not reflected by any of the performance measures in this study, participants' preferences for adaptable and hybrid automation were explained by a range of problems associated with adaptive automation: (1) the threshold for triggering changes in automation level was not always considered appropriate, (2) system-initiated changes can be confusing, (3) the system did not allow the operator to choose a different but reasonable path, and (4) the operator was not able to predict when a transition between automation levels was going to happen (Billings & Woods, 1994). These problems may lead to low acceptance, reduced trust, and poor performance. These disadvantages of adaptive automation may be associated with the specific invocation technique used in this experiment: a performance-based technique. To date, few studies have evaluated performance-based adaptation (Parasuraman, Cosenzo, & de Visser, 2009; Calhoun, Ward, & Rufll, 2011; Cosenzo, Chen, Reinerman-Jones, Barnes & Nicholson, 2010), although performance advantages were shown with this type of adaptive system, the entire adaptive cycle (invoking automation when performance is poor; invoking less automation for operator re-engagement when performance meet a threshold).

More participants preferred adaptable automation over the hybrid scheme. This may be due, in part, to the above disadvantages of adaptive automation which is part of the hybrid system. One possible solution to this problem could be the introduction of a management-by-consent strategy (Olson & Sarter, 2001) for the adaptive component of the hybrid system. In management-by-consent approach, the automation is not allowed to take action (in this case, change the level of an automation function such as changing from auto guidance to auto control) until explicitly operator consent has been received. Therefore, it gives the operator a stronger sense of control and predictability.

Operators' strategies for automation management in the adaptable and hybrid schemes were examined closely and can be summarized as follows. In the adaptable automation condition, operators decided and planned the use of an automation level based on perceived task difficulty (e.g. rotations were considered as more difficult than translational movement, multi-axis control more difficult than mono-axis hand control), confidence in his/her manual skills, and perceived benefits and costs of automation. For example, if an operator considered moving the arm across the table diagonally (multi-axis control) to be relatively difficult, he/she might plan to start with the medium level of trajectory automation, then use the higher level of automation for the second segment, and finally switched back to the medium level for last segment. On the other hand, when the operator had high confidence in his/her ability of manually operating the arm and was concerned about costs associated with failures of higher level automation, he/she was more likely to decide to use the medium level for entire scenario, including the difficult segment. If the operator was less confident in his/her manual skills, he/she may choose to use the high level at all times and plan to take over and perform the task manually only if the automation fails. After starting the movement of the arm, the operators tended to stick to their original plan. However, in the case of a critical event, such as the encounter of a hazard alert, or operator struggling with the ongoing task, the operator then re-evaluated the situation and decided whether and how to change the level of automation. Figure 5.6 shows the process of choosing and changing the level of automation in the adaptable condition.

In the hybrid condition, operators actually adopted strategies that turned the system either into adaptable automation (i.e., the operator set the level of automation and system-initiated changes were never triggered), adaptive automation (i.e., the operator did not set or change the system state and, if the system initiated any changes to the automation level, the operator did not intervene or override), or fixed automation (i.e., the operator flew the arm at the medium level at all times). Note, however, that during an earlier pilot study where participants encountered automation failures, more interventions by the operator and collaboration between the operator and the system were observed.

In these automation failure trials, the second segment defined in the configuration file was lower than it was supposed to be. As a result, if the trajectory was flown by the high level of trajectory control automation, the EE would go off the visible path and could get close to the top of the workbench and trigger a collision warning. If the hazard avoidance was in the medium level, the clearance between the EE and the workbench might or might not become small enough to trigger the switch to the high level of hazard avoidance would in the adaptive or hybrid condition. The arm might be able to continue moving from the first waypoint to the second waypoint with the EE being red throughout the segment. This strange situation could happen either when the operator set the trajectory automation to the high level, or if the operator flew along the visible path thus triggered the system-initiated change to the high level of trajectory control automation. Two participants performed this scenario with the hybrid scheme. They both experienced repeated automation failures.

The first set the trajectory automation to the high level at the end of the first segment because of difficulty in performing the rotation. The auto trajectory control then moved the arm to a warning zone. Thus the participant overrode the system and set the trajectory automation to the medium level and flew the arm manually along the visible path. However the system considered the EE to be off the path, thus took over. Again the automation a warning was triggered, and subsequently the operator overrode the trajectory automation. This repeated a third time until the EE reached the second waypoint. After the second waypoint, i.e. during the third segment, the participant flew the arm and everything went well. In this trial, the switches initiated by the system were annoying to the participant. One time when she overrode the automation level she said: ‘stopped it’.

The second participant performed the task with the hybrid automation scheme. He first set the level of trajectory automation to be high from the very beginning. Same as the first three participants, the automation brought the arm to a warning zone. The participant clicked on the button for the medium level of trajectory automation immediately after seeing a warning. Then he tried to fly the arm back to the path, but the auto took over again. The “high->warning->medium->high” cycle repeated two more times. The participant also changed the hazard avoidance automation from medium level to high level when he saw the warning for the first time.

As can be seen, the operator started the initial choice of automation level in the same way as in the adaptable scheme. During the scenario, when the system initiated a change, the operator would observe the system behavior and evaluate the validity of the change. If the system initiated change was not considered as appropriate (automation failure), the operator tended to override the system initiated change and adjust the automation level. The operator could still change the level of automation based on the initial plan or critical events at any point of the scenario. Figure 5.7 shows the process of operator’s automation management in the hybrid scheme. These frameworks are only based on observations on less than 20 participants. They provide an initial attempt in operator’s interaction with adaptable and hybrid automation systems, but would need to further validation using a larger scope of empirical data in the future.

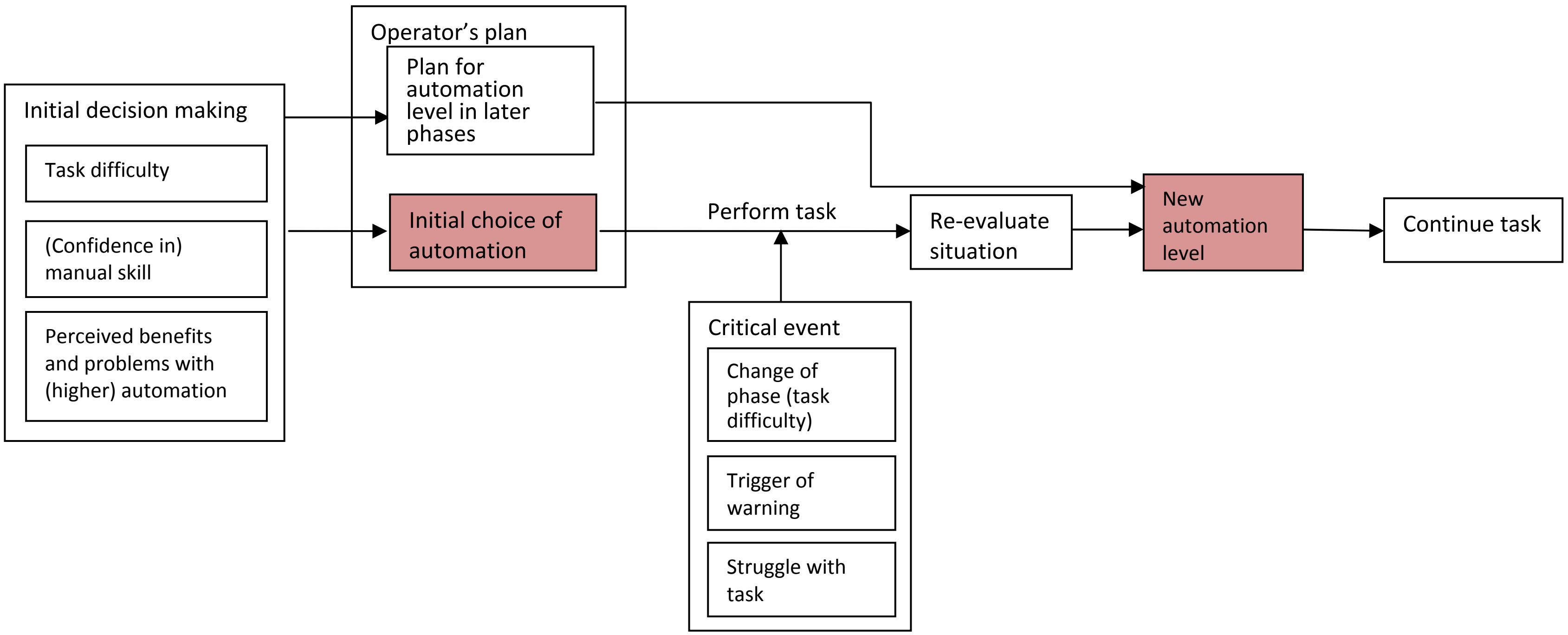


Figure 5.7: The process of LOA management in adaptable automation scheme (red boxes show operator initiated setting/change to the system state)

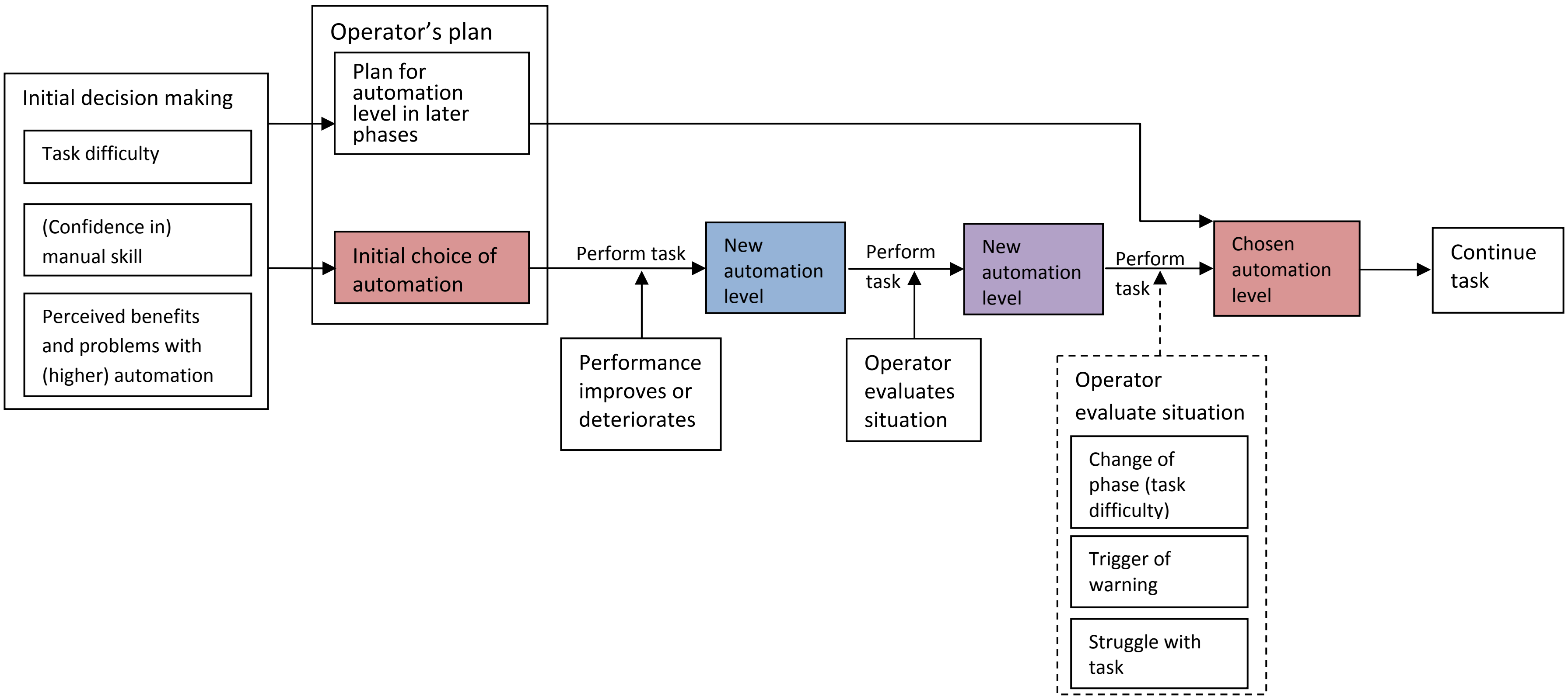


Figure 5.8: The process of LOA management in the hybrid scheme (red boxes show operator initiated settings/changes to the system state; blue boxes show system-initiated settings/changes; and purple boxes show the result of a collaboration between the operator and the system

Chapter 6

Conclusion

Operators in complex data-rich domains, such as aerospace operations, process control, and medicine, face considerable attentional demands and often need to make decisions and solve problems under time pressure and uncertainty. To support operators in meeting these demands, a wide range of automation technologies have been developed and introduced. However, the particular implementation and utilization of these technologies, as well as their coordination with human operators, involves its own set of challenges for designers.

In particular, the proper and timely allocation of tasks and functions to human operators and their automated systems is critical and needs to be based on a thorough analysis of automation properties and capabilities as well as performance-shaping factors, such as workload and task difficulty. Past research and operational experience have shown that fixed designs where tasks and functions are assigned on an a-priori basis often lead to unbalanced workload and poor overall system performance (Sarter & Woods, 1995; Wiener, 1988). A more context-sensitive approach to function allocation is needed instead, which can take two forms (Opperman, 1994): (a) adaptable (user-controlled) or (b) adaptive (system-controlled). A review of the literature, however, suggests that neither approach by itself is very effective. A hybrid approach may be needed that combines the benefits but avoids the disadvantages of either method. In addition to identifying the proper locus of control for task assignments, it is critical to design interfaces that support users in staying informed about, and overriding and adjusting, function allocations without distracting them from their primary tasks.

The main goals of this research are therefore to (1) systematically examine the performance effects of different stages and levels of automation during routine operations and in case of automation failures, and (2) implement and compare the effectiveness of three dynamic function allocation schemes: adaptive (system-controlled), adaptable (user-controlled), and a hybrid approach where the default mode was adaptive but operators were able to override system selections. The research was conducted in the context of space teleoperation, in particular operation of the robotic arm on the International Space Station.

First, we developed improved visual and tactile warnings to better support operators in monitoring the arm configuration and avoiding collisions and singularities. This first step, the development of improved stage 1 and 2 automation, was necessary because the current interface that is used by NASA makes it extremely difficult for astronauts (and for participants in our research) to perform their tasks. The benefit of the improved arm configuration GUI that highlighted the hazard zones is that it represents an evolutionary design that would be fairly easy for NASA to implement. It also requires minimal training for astronauts familiar with the original interface. Still, the integration of visual highlighting with the camera view and the tactile warnings, while requiring more significant changes to the existing operator station, would be an even better solution that can improve operators’ attention management and overall performance significantly.

Next, we conducted a study investigating the effects of different levels of automation for supporting all four stages of information processing which were represented by hazard avoidance, camera selection, and trajectory control. Results show that high levels of automation result in improved performance during routine scenarios but incur performance costs in cases of automation failure. This effect is especially pronounced with stage 4 automation – trajectory control. Performance in terms of completion time and trajectory deviations improved as the level of automation (LOA) increased during routine tasks, but it significantly deteriorated in automation failure scenarios. Camera selection (stage 3 automation) benefited somewhat from automation but this effect was limited because participants perceived this form of assistance to be less effective and thus adopted it to a much lesser extent than trajectory guidance and auto control. Finally, stage 1 & 2 automation (hazard avoidance) led to improved performance in routine scenarios, independent of the level of automation. However, as with trajectory control, a trend towards poor performance and reduced safety of operations was observed in case of failures of the hazard avoidance automation.

The mixed findings from the above study suggested the need for a more thorough review of earlier findings on the benefits and disadvantages of various stages and levels of automation. A meta-analysis of 17 studies related to this topic was conducted, which confirmed the tradeoff between costs and benefits of automation as a function of system reliability and performance.

Both the findings from the above study and meta-analysis, as well as operational experience with current automation technologies, highlight the need for context-sensitive use of automation. In chapter 5, we describe a study that comparatively evaluated three different dynamic function allocation schemes: adaptive, adaptable, and hybrid. The findings show that, overall, adaptable automation seems preferable over adaptive automation because of the increased sense of control on the part of the operator and the confusion caused by unexpected changes triggered by adaptive automation. Also, performance with adaptable automation was superior to that with adaptive automation, which can be explained, in part, by participants' prolonged use of higher levels of automation. This tendency is worrisome given the findings from our earlier study which showed the significant performance cost associated with failures of high levels of automation. The hybrid system, which combines adaptive and adaptable automation, seems to be a promising means of supporting both enhanced performance and high operator acceptance. However, its particular implementation still needs to be refined to eliminate problems associated with its adaptive component, such as unpredictable changes initiated by the system, lack of flexibility for operator to execute their own plan, and lack of perceived control.

In summary, this line of research contributes to a better understanding of human-automation collaboration and coordination and provides input to models of joint system performance. First, it systematically examined the effects of LOA for different stages of automation through both empirical study and a meta-analysis of the relevant literature. It extended previous studies by investigating multiple stages of automation in a single study and examining LOA impacts on various aspects including performance, workload, scan pattern, and subjective rating. The findings from this research provide important guidance for choosing appropriate levels of automation for various tasks and system designs. They highlight that the desirable level of automation depends on the stage of information processing it is supposed to support.

Another significant contribution of this work is that it represents the first effort to develop and empirical evaluate a hybrid function allocation scheme that combines both an adaptive and an adaptable component. Our findings highlight the potential for such a hybrid system to improve joint system performance and reduce the risk of confusion while ensuring acceptability to users by supporting a high level of perceived control. However, some challenges associated with adaptive automation still need to be addressed to make such a mixed approach even more successful. For example, employing a ‘management-by-consent’ technique (Olson & Sarter, 2000) where the automation cannot take any action without explicit operator agreement may be preferable as it provides operators with an even greater sense of control over system-initiated changes.

One important recommendation following from this research is that high levels of automation should be employed only in connection with high reliability technologies. For systems that are more likely to experience failures, lower levels of automation are preferable and may be sufficient to support successful joint system performance. This consideration is particularly important for stage 4 automation. Also, lower levels of automation may be preferable given that space teleoperation tends to be performed over extended periods of time. This implies that astronauts may experience vigilance decrements that can be exacerbated by the use of high levels of automation.

To improve the productivity and safety of space operations, which are currently labor intensive and costly, would require the development and introduction of a significant number and range of new automation technologies. The findings of this research provide empirically based guidelines that support designers in developing these new technologies. Specifically, the findings help designers and mission planners (a) anticipate and avoid potential problems in function allocation strategies in system design before new systems are introduced, and (b) ensure that these systems and their function allocation strategies can be implemented seamlessly and in a way to minimize transient or longer-term impacts on performance in space exploration missions. The findings of this dissertation also have implications for the design of automation schemes in other complex domains.

Several directions for future research are proposed. First, given the limitations associated with the camera selection automation in Chapter 3, more work on dynamic function allocation schemes in support of stage 3 information processing is needed. One example of a potential application is clinical decision support system in the medical domain.

Second, another important direction for future work is to examine in more detail, and build models of, operator strategies for choosing levels of automation in adaptable and hybrid systems. Contributing factors to LOA selection could be identified, and a quantitative model could then be developed to enable prediction of operator strategies in a given situation.

Third, it will be important to examine the effects of dynamic function allocation schemes on automation failure scenarios. Differences in the use of automation levels were found across the three dynamic function allocation schemes in Chapter 5. These differences accounted for most of the differences in routine performance. At the same time, the impact of automation level on failure performance with a fixed automation scheme was seen in the meta-analysis in Chapter 4. We expect that the use of automation level also has a significant impact on failure performance with dynamic function allocation.

Finally, the limitations of the final experiment described in Chapter 5 suggest that an increase in task difficulty and scenario complexity may be necessary to encourage more changes to responsibility assignments (both by the human operator and by the system) and thus allow for more extensive comparisons between function allocation schemes. Complex scenarios that could potentially be used for future experiments include grapple tasks (using end effector to grab a payload) and tracing tasks (using end effector camera to follow a specific path while maintaining the camera view at a certain angle), both of which involve more complicated and less intuitive combinations of coordinate frames. The complexity of these scenarios may create challenges for operators' attention management, and thus could induce a loss of system awareness in adaptive automation and increased workload in adaptable automation due to the additional automation and interface management task. With more complex tasks, the hybrid function allocation scheme would be expected to be particularly beneficial.

The efforts described here, and future efforts in investigating the design of function allocation schemes show promise for better design of human-automation systems, and consequently improved joint-system performance. Through continued efforts, we can minimize the problems associated with current automation systems, maximize the benefits of automated aids, and finally increase the safety and efficiency of operations in a wide range of domains through improved automation design.

Appendices

Appendix I: Coding Sheet for Meta-Analysis

Coding sheet for meta-analysis for automation stages and levels

<table>
<tr><td>Author</td><td></td></tr>
<tr><td>Title</td><td></td></tr>
<tr><td>Publication</td><td></td></tr>
<tr><td>Domain</td><td></td></tr>
<tr><td colspan="2">Factors</td></tr>
<tr><td>1) Stages & levels</td><td>
Condition A:
<table>
<tr><td></td><td>S1</td><td>S2</td><td>S3</td><td>S4</td></tr>
<tr><td>None</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Low</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Medium</td><td></td><td></td><td></td><td></td></tr>
<tr><td>High</td><td></td><td></td><td></td><td></td></tr>
</table>
Name:
level in 10-point scale
Condition B:
<table>
<tr><td></td><td>S1</td><td>S2</td><td>S3</td><td>S4</td></tr>
<tr><td>None</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Low</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Medium</td><td></td><td></td><td></td><td></td></tr>
<tr><td>High</td><td></td><td></td><td></td><td></td></tr>
</table>
Name:
level in 10-point scale
Condition …:
</td></tr>
<tr><td>Summary</td><td>
Stages and levels
<table>
<tr><td></td><td>S1</td><td>S2</td><td>S3</td><td>S4</td></tr>
<tr><td>None</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Low</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Medium</td><td></td><td></td><td></td><td></td></tr>
<tr><td>High</td><td></td><td></td><td></td><td></td></tr>
</table>
</td></tr>
<tr><td>2) Adaptive</td><td></td></tr>
<tr><td>3) Trust</td><td></td></tr>
<tr><td>4) Reliability</td><td></td></tr>
<tr><td rowspan="4">Dependent Measures</td><td>Primary task performance:</td></tr>
<tr><td>Perceived workload</td></tr>
<tr><td>Secondary task performance:</td></tr>
<tr><td>SA</td></tr>
</table>

	Subjective rating
	Physiological measurement
Results	
Primary task	Effect of DOA F(,) =, p=
Secondary task	F(,)= , p=.
Situation awareness	F(,)= , p=.
Mental workload	
Subjective rating	
Physiological measurement	

Appendix II: Cognitive Task Analysis

Critical decisions made by BORIS operators

Operating the robotic arm in BORIS includes four major steps: Camera, Frame, Rate and Mode. Critical decisions involved in these steps are described below. Although the CFRM strategy is used by the operator before starting operating the arm with hand controllers, there are some remaining decisions about C, R, and M during the operation.

Cameras

Before moving the arm

1. Choose which 2-3 cameras to use for each of the three purposes (big picture, task view, clearance view), and how to observe the end point of the target location in camera view (based on the result of Triangulation, which also helps with camera choosing)
2. Decide the orientation of the cameras.
3. Zoom the camera properly
4. Most of the time, the window view is kept in the largest monitor. In situations when more than 2 camera views are needed, the operator need to decide how to use 'multiplex' (2 views on one monitor).

During operation

5. Decide when to pan/tilt the cameras (e.g. to have a good view on clearance)
6. Decide when to switch to other cameras or add an additional camera view
7. Decide whether and when to ask M2 for help

Frames

Before moving the arm

1. Decide which frames to use (FOR (1-4), display frame: forward-base, port-base, etc., command frame: internal, external)
2. Develop a plan for controls of all DOFs and think about expected movements of the arm

During operation

Nothing (changing frame during operation is not encouraged).

Rates

Before moving the arm

1. Choose coarse or vernier rate according to current position (clearance consideration, distance to target location) and the expected movement

During operation

2. Adjust the rate

If the end effector or the arm is within 1.5m to the structure, switch to or maintain at vernier.

If a large movement is required and not clearance consideration, switch to coarse.

Modes

Before moving the arm

Decide which mode to use: manual or auto (FOR or Joint)

During operation

Manual: Decide whether it is necessary to turn on the brakes when there is a pause or an interruption

Auto: Decide when there is a potential problem and need to pause

Besides the four steps described above, some others strategies are also important during operating the arm.

Call-out points

During grapple and berthing, the operator needs to watch the height of the end effector (or payload) and call out at 1.5m above the pin/guides, 1m above, 0.5m above, 1/3 pin height (or 1 guide height), right above.

Scan

The operator also needs to decide when and where to look at so that the whole display is well scanned.

Appendix III: Training Checklist

1. Experiment Overview
    - ☐ Explain BORIS
    - ☐ Virtual Environment (Name of 4 walls, arm, boxes, table, payload)
2. Robotic Kinematics
    - ☐ Read Degree of Freedom (Page 1)
        - i. Explain DOF using the robotic arm model (shoulder 2, elbow 1, wrist 3)
    - ☐ Euler Sequence / Right Hand Rule
        - ii. Explain Euler Sequence and RHR using the block/EE model
    - ☐ Joints of the Arm
        - iii. Explain names of the joints
        - iv. Explain work envelope using the robotic arm model
        - v. Operational Work Envelope (explain use robotic arm model)
            1. Joint limit (Page 14 of Handbook)
            2. Arm Singularity (Page 15 of Handbook)
            3. Self-Collision (Page 1)
    - ☐ Euler Sequence Practice (slide 7)
3. Frames
    - ☐ Read DYF (explain using the block)
    - ☐ Read FOR (explain using the EE model)
    - ☐ CDF: relationship b/w the hand controller & movement of the EE
        - i. Internal (explain using EE model)
        - ii. External (explain using both)
    - ☐ Hand Controller Frame
        - iii. Demonstrate to the participant using the joystick
        - iv. Familiarize the participant with the controllers (distinguish between translational and rotational control)
4. Start BORIS
    - ☐ Joystick (Familiarize the participant with XYZ and PYR)
    - ☐ Explain Camera, Rate (slide 13, 14)
    - ☐ Let participant move the arm using the hand controller
5. Operation Procedures
    - ☐ Review 3 Flight Rules (slide 11)
    - ☐ Subtasks in Sequence (CFRM)
        - i. Camera
        - ii. Frames
        - iii. Rate
        - iv. Mode
        - v. Operating the arm
    - ☐ Tolerance
6. Perform Practice Task
    - ☐ Explain the Virtual Environment with the floor plan of the room
    - ☐ Explain how to select cameras using floor plan
    - ☐ Explain the view of each camera (Clearance, Task, Big picture)
    - ☐ Ask the participant about his/her plan for moving the arm

Appendix IV: Training Review Checklist

1. Euler's sequence (XYZ, PYR)
2. Work Envelope (Mechanical, Operational)
3. 3 Singularities
4. Self-collisions
5. Frames
   a. Display Frame
   b. Frame of Resolution
   c. Command Frame (Internal, External)
6. Hand Controller Frame
7. Subtasks in Sequence (CFRM)
8. 3 Flight Rules
9. Reach limits for each joint
10. Camera views (Big picture, task, clearance), Rate
11. Tolerance
12. Spoking scan
13. Multi-axis hand control

Appendix V: Background Data Questionnaire

1. Age: ___________________

2. Gender: _______________

3. Are you: right-handed left-handed ambidextrous

4. Do you wear glasses or contact lenses? ____________
If so, which of the two?_____________________________

5. What department are you in?__________________ Do you have an area of specialization?
____

6. Are you an undergraduate (year_________), Master's or Ph.D. student?

7. Please rate your current level of fatigue (circle one)

8. Do you have any experience with Virtual 3-D environments (e.g. 3-D games, CAD, 3-D graphic design, etc.)? If so, can you please describe this experience?

9. Do you have any experience with joysticks or game controllers (e.g. computer games, video games, robotic manipulation)? If so, can you please describe this experience?

10. Have you ever played video games? Yes. No.

11. Do you currently play video games? Yes. No.

12. How long have you been playing video games?
a. 6 months
b. 1 year
c. 2-5 years
d. 5-10 years
e. 10 or more years

13. How often (approximately) do you play video games?
a. daily d. once in 6 months
b. weekly e. once a year
c. once a month f. less than once a year or never

14. In your opinion, how good are you at playing video games?
a. highly skilled
b. fairly skilled
c. not very skilled
d. not skilled at all

16. What kind of video/computer games do you play the most?
a. first person
b. role-playing/strategy
c. arcade/fighting (please specify: 2D   3D)
d. simulation (driving, flying, etc.)
e. sports (please specify ________________________________)
f. other ______________________

Do you have any prior knowledge about robotic arm control operations?
(Yes   No)

If so, can you please explain the kind and level of experience?

Appendix VI: Debriefing Questionnaire for Chapter 2

1. Please rate your current level of fatigue (circle one)

2. Please rate your overall performance on the last task.

1	2	3	4	5	6	7	8	9	10
very poor								perfect	

3. Were there tasks (or aspects of tasks) that seemed particularly difficult? If so, please explain.

4. Please describe the strategies you adopted to control the arm, e.g. how you chose the frames of references, how you decided the hand motion.

5. Please describe the strategies you adopted to monitor the visual cues and graphic user interfaces. For example, did you focus on one or several of the windows? Did you focus more on the visual cues or did you scan all the displays frequently?

6. Under which automation scheme (scheme 1-4) was it easiest to perform the following tasks:

Choosing camera views: 1 2 3 4
Choosing frames of references: 1 2 3 4
Monitoring the arm: 1 2 3 4
Monitoring the clearance: 1 2 3 4
Monitoring the system: 1 2 3 4
Overall, which automation scheme do you like best? Please explain.

7. Which interface (1-3) do you prefer for the following tasks:

Choosing camera views: 1 2 3
Choosing frames of references: 1 2 3
Monitoring the arm: 1 2 3
Monitoring the clearance: 1 2 3
Monitoring the system: 1 2 3
8. Overall, which automation scheme do you like best? Please explain.

9. What was the most frustrating aspect of the experiment?

10. Do you have any suggestions for how we might improve upon our current interface design?

11. Were there any issues with the current simulator setup? (anything out of place, awkward timing, etc.)?

12. In your opinion, was the training adequate for performing your tasks in the experiment?

1 2 3 4 5 6 7 8 9 10
Inadequate Excellent

If you felt that the training was less than adequate, please explain what task(s) you did not feel sufficiently prepared for.

Appendix VII: Tasks for Chapter 2

(Starting points: end effector above the black berth; Pitch, Yaw, Roll= -90, 0, 0)

Practice
X, Y, Z = 900, 200, -300
Pitch, Yaw, Roll = -90, -45, -90
FOR: end effector
Display frame: base forward
Command frame: student choice

Task 1
X, Y, Z = 900, -900, -200
Pitch, Yaw, Roll = -90, 0, 0
FOR: end effector
Display frame: base forward
Command frame: student choice

Task 2
X, Y, Z = 100, -900, -100
Pitch, Yaw, Roll = 0, 90, 0
FOR: end effector
Display frame: base forward
Command frame: student choice

Task 3
X, Y, Z = 800, 400, 0
Pitch, Yaw, Roll = 0, -90, 0
FOR: end effector
Display frame: base forward
Command frame: student choice

Appendix VIII: Debriefing Questionnaire for Chapter 3

Function Allocation in Complex Systems

Debriefing Questionnaire

1. Please describe the strategies you adopted to perform the following tasks. Did you encounter any challenges? How did you solve them?

1.1 How did you go about planning a trajectory?

1.2 While flying the arm, how did you decide the hand motion (e.g. which way to move the hand, whether to move along one axis or multiple axes)?

1.3 How did you monitor the arm as you flew a trajectory? Did the robotic arm move as expected (always, most of the time, sometimes, rarely, never)?

1.4 How did you know (or estimate) when to turn when executing the trajectory?

1.5 How did you know (or judge) that you had reached the end point?

2. Please describe the strategies you adopted to select cameras. How much did you depend on the camera recommendations when they were offered?

3. Please describe the strategies you adopted to monitor multiple windows of visual cues and graphic user interfaces. For example, did you focus on one or several of the windows? Did you focus more on the visual cues or did you scan all the displays frequently?

4. Were there tasks (or aspects of tasks) that seemed particularly difficult? If so, please explain.

What was the most frustrating aspect of the experiment?

5. Which one(s) of the different kinds of automation (camera selection, hazard avoidance, trajectory control) were helpful for these sub-tasks? If you preferred certain types of automation for certain subtasks, please identify and explain.

Choosing camera views:

Monitoring the arm configuration:

Monitoring the clearance:

Monitoring the movement of the arm:

Fly the arm:

For hazard avoidance, which level would you prefer? Please explain. Alert only. alert and automatic stop.

6. Do you have any suggestions for how we might improve upon our current system (e.g. system functions, automated aids, interface design)?

7. Were there any issues with the current simulator setup? (anything out of place, awkward timing, etc.)

8. In your opinion, was the training adequate for performing your tasks in the experiment?

1 2 3 4 5 6 7 8 9 10

Inadequate Excellent

If you felt that the training was less than adequate, please explain what task(s) you did not feel sufficiently prepared for.

Appendix IX: Debriefing Questionnaire after Individual Scenarios for Chapter 3

Debriefing Questionnaire for Individual Scenarios

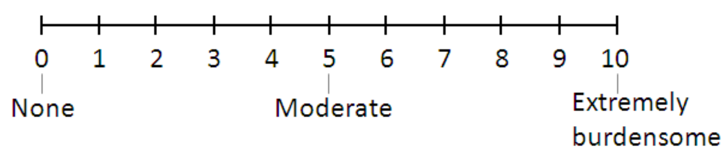
1. Have you completed the scenario/fly-to task (i.e. move the EE to the target location and orientation within the 'tolerance')?

2. How confident are you about the completion of the scenario?

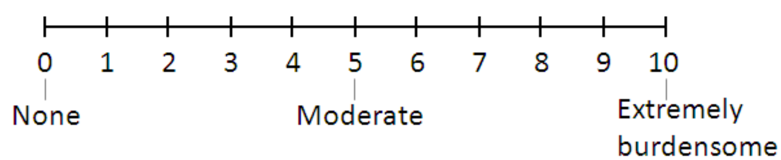
1 2 3 4 5 6 7 8 9 10

not confident ... fully confident

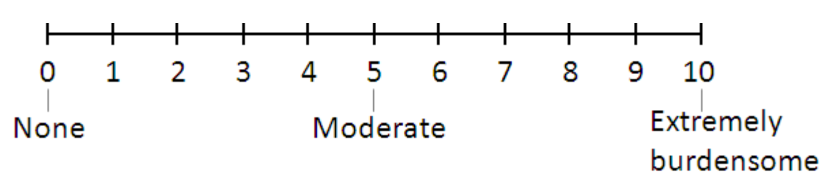
3. What level of workload on average did you experience during this scenario? What is the maximum and minimum workload and when did they appear?



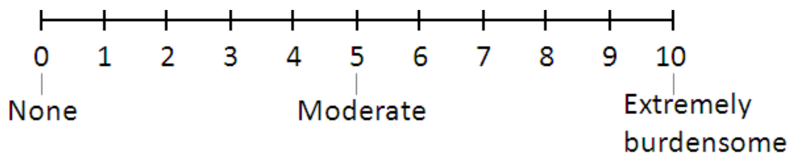
4. Please rate your average, maximum and minimum visual workload during this scenario.



5. Please rate your average, maximum and minimum cognitive workload during this scenario.



6. Please rate your average, maximum and minimum motor workload during this scenario



7. Please rate your overall performance during the scenario:

1 2 3 4 5 6 7 8 9 10

very poor perfect

8. Did you notice any warning during the scenario? If yes, when did you notice each warning and what problem did you think you encountered?

Appendix X: Debriefing Questionnaire for Chapter 5

Function Allocation in Complex Systems

Debriefing Questionnaire

1. Please describe the strategies you adopted to perform the following tasks. As part of your answers, please indicate whether you encountered any challenges and how you resolved them.

1.1 How did you go about planning a trajectory?

1.2 While flying the arm, how did you determine the appropriate control inputs (e.g., which way to move the controller, whether to move along one axis or multiple axes at the same time)? Please also comment on how difficult it was to figure out which control input would result in the desired path.

Did the robotic arm move as expected? (Please circle the proper description.)

never rarely sometimes most of the time always

1.3 How did you monitor the arm while flying a trajectory?

1.4 How did you determine that you had reached the end point of each of the three legs?

2. Please describe the strategy you adopted for monitoring the various camera views and the graphic user interface on the left monitor. For example, did you develop a standardized scan or was your monitoring driven by the phase of flight and/or events/cues that appeared on either monitor? Which display elements did you monitor the most?

3. Were there tasks (or aspects of tasks) that seemed particularly difficult? If so, please explain.

4. Please comment on the adaptive automation scheme (where the automation level changed on its own): Did you agree with the changes in automation level?

Did you ever miss when the automation level had changed?

Was it confusing that the automation level changed on its own?

Other comments:

5. Please comment on the adaptable automation scheme:

Did you think it is necessary/desirable that you have the option to manually change/override the level of automation?

Was it easy or difficult to make changes to the automation level? Please explain.

Other comments:

5. Please comment on the hybrid automation scheme:

Did you think it is necessary/desirable that you have the option to collaborate with the system, e.g. manually change/override the level of automation?

Was it easy or difficult to make changes to the automation level or collaborate with the system? Please explain.

Other comments:

6. Please describe the strategies you used to keep track of the system state (i.e., the levels of hazard avoidance and of trajectory control automation) in the adaptive and hybrid automation schemes.

Did you find the system state GUI helpful? Do you have any suggestions for improvements?

7. Overall, which automation scheme (fixed, adaptive, adaptable or hybrid) did you like best? Please explain.

8. Do you have any suggestions for how we might improve upon our current system (e.g. system functions, automated aids, and interface design)?

9. Were there any issues with the current simulator setup? (anything out of place, awkward timing, etc.)

10. In your opinion, was the training adequate for performing your tasks in the experiment?

1	2	3	4	5	6	7	8	9	10
Inadequate									Excellent

If you felt that the training was less than adequate, please explain what task(s) you did not feel sufficiently prepared for.

Appendix XI: Debriefing Questionnaire after Individual Scenarios for Chapter 5

Debriefing Questionnaire for Individual Scenarios

1. Have you completed the scenario/fly-to task (i.e. move the EE to the target location and orientation within the ‘tolerance’)?

   O yes O no

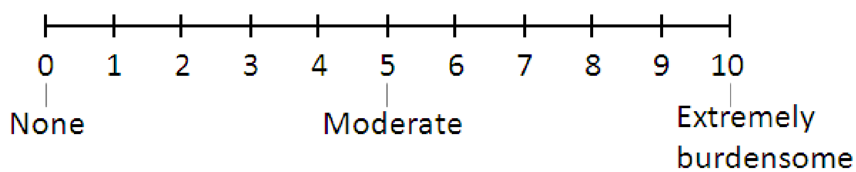
2. How confident are you about the completion of the scenario?

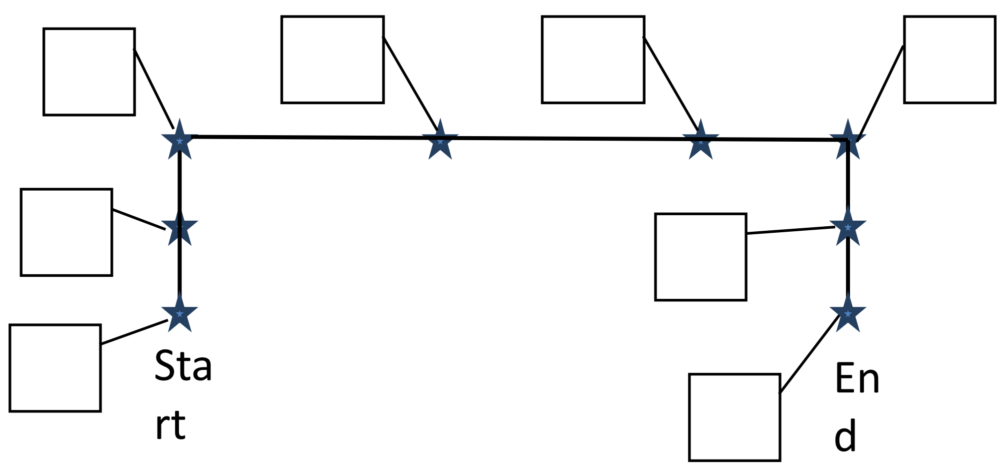
1 2 3 4 5 6 7 8 9 10

not confident ... fully confident

3. What level of overall workload did you experience at each point during this scenario? Please rate your workload on a 0-10 scale.

   Scale for rating workload





4. What level of visual workload/monitoring demands on average did you experience during this scenario? What is the maximum and minimum workload and when did they appear? Please rate your workload on a 0-10 scale. Please indicate when/where you experienced the maximum and minimum workload on the figure below.

   Average workload:

Maximum workload:
When/where experienced:
Explain:

Minimum workload:
When/where experienced:
Explain:

5. Please rate your average, maximum and minimum cognitive workload during this scenario.

Average workload:

Maximum workload:
When/where experienced:
Explain:

Minimum workload:
When/where experienced:
Explain:

6. Please rate your average, maximum and minimum motor workload (for controlling the arm) during this scenario.

Average workload:

Maximum workload:
When/where experienced:
Explain:

Minimum workload:
When/where experienced:
Explain:

7. Please rate your overall performance during the scenario:

1 2 3 4 5 6 7 8 9 10

very poor perfect

8. Did you notice any warnings during the scenario? If yes, when did you notice each warning and what problem did you think you encountered?

9. Please recall as many as you can, the changes to the system state during this scenario.

For each change, please recall

- the level of hazard avoidance automation,
- the level of trajectory control automation, and
- who (you or the system) made the last change.

10. Please rate your ability of tracking the system state during the scenario:

1 2 3 4 5 6 7 8 9 10

very poor perfect (can track all the changes)

11. Please describe any difficulties you may have experienced and how you tried to monitor system state.

References

Abbott, K., Slotte, S. M., & Stimson, D. K. (1996). *The interface between flightcrews and modern flight deck systems*. Washington, DC: Federal Aviation Administration.

Arciszewski, H. F. R., Greef, T, E. de, & Delft, J. H. van. (2009). Adaptive automation in a naval combat management system. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, 39*(6), 1188-1199.

Bahner, E.J., Huper, A. D., & Manzey, D. (2008). Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies, 66* (9), 688–699.

Bailey, N. R., Scerbo, M.W., Freeman, F.G., Mikulka, P.J., & Scott, L. A. (2003). The effects of a brain-based adaptive automation system on situation awareness: The role of complacency potential. In *Proceedings of the Human Factors & Ergonomics Society 47th Annual Meeting* (pp. 1048-1052). Santa Monica, CA: Human Factors & Ergonomics Society.

Bailey, N. R., Scerbo, M.W., Freeman, F. G., Mikulka, P. J., & Scott, L. A. (2006). Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Human Factors, 48* (4), 693–709.

Bainbridge, L. (1983). Ironies of Automation. *Automatica, 19,* 776-779*.*

Barnes, M., & Grossman, J. (1985). *The intelligent assistant for electronic warfare systems* (Rep. No. NWC TP 5885) China Lake, CA: U.S. Naval Weapons Center.

Billings, C. E., & Woods, D. D. (1994). *Concerns about adaptive automation in aviation systems. In M. Mouloua & R. Parasuraman (Eds.), Human performance in automated systems: Current research and trends (pp. 24–29). Hillsdale, NJ: Erlbaum.*

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

Barnes, M., Parasuraman, R.,&Cosenzo, K. (2006). Adaptive automation for military robotic systems. In *RTO-TR-HFM-078 Uninhabited military vehicles: Human factors issues in augmenting the force* (NATO Tech. Rep., pp. 420–440). Brussels: NATO Research and Technology Organization.

Brandan, M. A. M. (2007) Influence of spatial orientation and spatial visualization abilities on space teleoperation performance. M.S. theis at Massachusetts Institute of Technology.

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology,42*, 249–268.

Caleefato, C., Montanari, R. Tango, F. (2007). Advanced Drivers Assistant Systems in Automation.

Calhoun, G. L., Draper, M. K., & Ruff, H. A. (2009). Effect of Level of Automation on Unmanned Aerial Vehicle Routing Task. In *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting* (pp. 197-201). Santa Monica, CA: Human Factors and Ergonomics Society.

Calhoun, G. L., Ward, V.B.R., & Ruff, H. A. (2011). Performance-based adaptive automation for supervisory control. In *Proceedings of the Human Factors & Ergonomics Society 55th Annual Meeting* (pp. 2059-2063). Santa Monica, CA: Human Factors & Ergonomics Society.

Calhoun, G., L., Ruff, H. A., Spriggs, S., Murray, C. (2012). Tailored performance-based adaptive levels of automation. In *Proceedings of the Human Factors & Ergonomics Society 56th Annual Meeting* (pp. 413-417). Santa Monica, CA: Human Factors & Ergonomics Society.

Chen, H., Santos, J., Graves, M., Kim, K., & Tan, H. Z. (2008). Tactors Localization at the Wrist. In Ferre (Ed.) Haptics: Perception, Devices and Scenarios: Proceedings of 6th International Conference, EuroHaptics, 209-218.

Cholewiak, R. W., & Collins, A. A. (2000). The generation of vibrotactile patterns on a linear array: Influences of body site, time and presentation mode. Perception and psychophysics, 62, 1220-1235.

Cholewiak, R. W., Brill, J. C., & Schwab, A. (2004). Vibrotactile localization on the abdomen; Effects of place and space. Perception and Psychophysics, 66, 970-987

Currie, N. J. & Peacock, B. (2002). International space station robotic systems operations – A human factors perspective. Proceedings of Human Factors and Ergonomics Society 46th Annual Meeting, 26-30.

Clamann, M. P., & Kaber, D. B. (2003). Authority in Adaptive Automation Applied to Various Stages of Human-Machine System Information Processing. *In Proceedings of Human Factors and Ergonomics Society 47th Annual Meeting, Cognitive Engineering and Decision Making*, (pp. 543-547). Santa Monica, CA: Human Factors & Ergonomics Society.

Clamann, M. P., Wright, M. C., & Kaber, D. B. (2002). Comparison of Performance Effects of Adaptive Automation Applied to Various Stages of Human-Machine System Information Processing. Human Factors and Ergonomics Society 46th Annual Meeting Proceedings, Cognitive Engineering and Decision Making, 342-346.

Cook, R. I., & Woods, D. D. (1994) Operating at the sharp end: The complexity of human error. In M.S. Bogner (ed.), Human Error in Medicine, 255-301.

Cosenzo, K., Chen, J., Reinerman-Jones, L., Barnes, M., & Nicholson, D. (2010, September). Adaptive automation effects on operator performance during a reconnaissance mission with an unmanned ground vehicle. In Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting (pp. 2135-2139). Santa Monica, CA: Human Factors and Ergonomics Society.

Crocoll, W. M., & Coury, B. G. (1990). Status or recommendation: selecting the type of information for decision aiding. In *Proceedings of the Human Factors Society 34th Annual Meeting* (pp. 1524-1528). Santa Monica, CA: Human Factors and Ergonomics Society.

Cummings, M.L., & Mitchell, P.J. (2007). Operator scheduling strategies in supervisory control of multiple UAVs. *Aerospace Science and Technology*, *11*(4), 339-348.

Currie, N. J. & Peacock, B. (2004). International space station robotic systems operations – A human factors perspective.

Degani, A., Shafto, M., & Kirlk A. (1996). Modes in automated cockpits: problems, Data analysis, and a modeling framework. In Proceedings of the 36th Israel Annual Conference on Aerospace Sciences.

Dekker, S. W. A., & Woods, D. D. (2002). MABA-MABA or abradacabra? Progress on human-automation coordination. *Cognition, Technology, and Work, 4,* 240-244.

de Visser, E. J., LeGoullon, M., Freedy, A., Freedy, E., Weltman, G., & Parasuraman, R. (2008). Designing an adaptive automation system for human supervision of unmanned vehicles: A bridge from theory to practice. In *Proceedings of the Human Factors & Ergonomics Society 52th Annual Meeting* (pp. 221-225). Santa Monica, CA: Human Factors & Ergonomics Society.

Di Nocera, F., Lorenz, B., & Parasuraman, R. (2005). Consequences of shifting from one level of automation to another: main effects and their stability. Human factors in design, safety, and management, 363-376.

Dixon, S. R., & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle flight control: a reliance-compliance model of automation dependence in high workload. *Human Factors*, *48* (3), 474-486.

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

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

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors, 37*, 387-394.

Erol, K., Hendler, J., & Nau, D. (1994). UMCP: Asound and complete procedure for hierarchical task network planning. In K. Hammond (Ed.), *AI planning systems: Proceedings of the 2nd International Conference* (pp. 249–254). Menlo Park, CA: AAAI Press.

Fadden, S., Ververs, P., & Wickens, C.D. (1998). Costs and benefits of head-up display use: A meta-analytic approach. Proceedings of the 42nd Annual Meeting of the Human Factors & Ergonomics Society. Santa Monica, CA: Human Factors Society.

Ferris, T. & Sarter, N. (2009). Supporting Anesthetic Monitoring through Tactile Display of Physiological Parameters. Proceedings of Human Factors and Ergonomics Society 53th Annual Meeting, 1664-1668.

Fitts, P. M (Ed.). (1951). *Human engineering for an effective air navigation and traffic control system*. Washington, DC: National Research Council.

Gacy, A. M., Wickens, C. D., Sebok, A., Gore, B. F., & Hooey, B. L. (2011). Modeling Operator Performance and Cognition in Robotic Missions. In Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting (pp. 861-865). Santa Monica, CA: Human Factors and Ergonomics Society.

Galster, S., & Parasuraman, R. (2001). Evaluation of countermeasures for performance decrements due to automated-related complacency in IFR-rated general aviation pilots. Proceedings of the International Symposium on Aviation Psychology. (pp. 245–249). Columbus: Ohio State University.

Gilliland, K., Schlegel, R. E., Shehab, R. L. (2001). Remote manipulator system operator training and performance metrics. Proceedings of the Bioastronautics Investigators Workshop, Abstract #031P.

Hayley D. (2002) Cognitive Task Analysis: Current Research. Presentation at 16.459 Humans & Automation Seminar- Spring 02 at Massachusetts Institute of Technology. http://web.mit.edu/16.459/www/

Hancock, P. A., & Scallen, S. F. (1996). The future of function allocation. Ergonomics in Design, 4(4), 24-29.

Hilburn, B., Jorna, P. G., Byrne, E. A., & Parasuraman, R. (1997). The effect of adaptive air traffic control (ATC) decision aiding on controller mental workload. In M. Mouloua & J. Koonce (Eds.), *Human-automation interaction: Research and practice* (pp. 84–91). Mahwah, NJ: Erlbaum.

Ho, C., Tan, H.Z. & Spence, C. (2005). Using spatial vibrotactile cues to direct visual attention in driving scenes. Transp. Res. Part F, 8, 397-412.

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors, 48*(1), 196-205.

Inagaki, T. (2003). Adaptive automation: Sharing and trading of control. In E. Hollnagel (Ed.), *Handbook of cognitive task design* (pp. 46–89). Mahwah, NJ: Lawrence Erlbaum.

Janis, I. L. (1982). Decision making under stress. In L. Goldberger & S. Breznitz (eds.), Handbook of Stress: Theoretical and Clinical Aspects (pp. 69-87). New York: Free Press.

Jones, L. A., Lockyer, B., & Piateski, E. (2006). Tactile display and vibrotactile recognition on the torso. Advanced Robotics, 20, 1359-1374.

Kaber, D. B., & Endsley, M. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science, 5*, 113–153.

Kaber, D.B., Onal, E., & Endsley, M.R. (2000). Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. Human Factors and Ergonomics in Manufacturing, 10 (4), 409–430.

Kaber,D.B., Prinzel,L.J., Wright,P., & Clamann, M.P. (2002). Workload-Matched Adaptive Automation Support of Air Traffic Controller Information Processing Stages. Technical report, NASA/TP-2002-211932.

Kaber, D.B., & Riley, J., (1999). Adaptive automation of a dynamic control task based on secondary task workload measurement. *International Journal of Cognitive Ergonomics 3* (3), 169–187.

Kaber, D. B., Riley J. M., Tan, Kl, Endsley, M. R. (2001). On the design of adaptive automation for complex systems. *International Journal of Cognitive Ergonomics, 5*(1), 37-57.

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

Kidwell, B., Calhoun, G. L., Ruff, h., A., & Parasuraman, R. (2012). Adaptable and adaptive automation for supervisory control of multiple autonomous vehicles. In *Proceedings of the Human Factors & Ergonomics Society 56th Annual Meeting* (pp. 428-432). Santa Monica, CA: Human Factors & Ergonomics Society.

Kirlik, A. (1993). Modeling strategic behavior in human automation interaction: Why an 'aid' can (and should) go unused. Human Factors, 35, 221-242.

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. International Journal of Human-Computer Studies, 40, 153–184.

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.

Lerch, F. J., Prietula, M. J., & Kulik, C. T. (1997, May). The Turing effect: The nature of trust in expert systems advice. In Expertise in context (pp. 417-448). MIT Press.

Lin, C. J., Yenn, T., & Yang C. (2010). Automation design in advanced control rooms of the modernized nuclear power plants. *Safety Science*. *48* (1), 63-71.

Lorenz, B., Di Nocera, F., Röttger, S., & Parasuraman, R. (2001) The Effects of Level of Automation on the Out-Of-The-Loop Unfamiliarity in a Complex Dynamic Fault-Management Task During Simulated Spaceflight Operations. In *Proceedings of Human Factors and Ergonomics Society Annual Meeting Proceedings* (pp. 44-48). Santa Monica, CA: Human Factors and Ergonomics Society.

Lorenz,B.,DiNocera,F.,Roettger,S., &Parasuraman,R. (2002). Automated fault management in a simulated space flight micro-world. Aviation, Space, & Environmental Medicine, 73, 886–897.

MacLean, K. E. & Hayward, V. (2008). Do It Yourself Haptics, Part II. IEEE Robotics and Automation Magazine, 15 (1), 104-119

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: an integrative review. Theoretical Issues in Ergonomics Science, 8(4), 277-301.

Manzey, D., Reichenbach, J., & Onnasch, L. (2008). Performance consequences of automated aids in process control: The impact of function allocation. In *Proceedings of Human Factors and Ergonomics Society 52nd Annual Meeting* (pp. 297-301). Santa Monica, CA: Human Factors and Ergonomics Society.

Manzey, D., Reichenbach, J., & Onnasch, L. (2009). Human performance consequences of automated decision aids in the states of fatigue. In *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting* (pp. 329-333). Santa Monica, CA: Human Factors and Ergonomics Society.

Manzey, D., Reichenbach, J., & Onnasch, L. (2012). Human Performance Consequences of Automated Decision Aids The Impact of Degree of Automation and System Experience. Journal of Cognitive Engineering and Decision Making, 6(1), 57-87.

Masalonis, A. J., & Parasuraman, R. (2003, October). Effects of situation-specific reliability on trust and usage of automated air traffic control decision aids. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 47, No. 3, pp. 533-537). SAGE Publications.

McFadden, S.M., Vimalachandran, A., & Blackmore, E. (2004). Factors affecting performance on a target monitoring task employing an automatic tracker. *Ergonomics*, *47* (3), 257–281.

Metzger, U., & Parasuraman, R. (2005). Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. *Human Factors*, *47* (1), 35–49.

Meyer, J., Feinshreiber, L., & Parmet, Y. (2003). Levels of automation in a simulated failure detection task. In *Proceedings of 2003 IEEE International Conference on Systems, Man and Cybernetics* (pp. 2101 – 2106). New York, NY: IEEE.

Miller, C., Goldman, R., Funk, H., Wu, P., & Pate, B. (2004). A playbook approach to variable autonomy control: Application for control of multiple, heterogeneous unmanned air vehicles. In *Proceedings of FORUM 60, the Annual Meeting of the American Helicopter Society* (pp. 2146–2157). Alexandria, VA: American Helicopter Society.

Miller, C., Funk, H., Goldman, R., Meisner, J., & Wu, P. (2005, July 22-27). Implications of Adaptive vs. Adaptable UIs on Decision Making: Why "Automated Adaptiveness" is Not Always the Right Answer. Proceedings of the 1st International Conference on Augmented Cognition, Las Vegas, NV.

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

Miller, C., Pelican, M., & Goldman, R. (2000). "Tasking" interfaces to keep the operator in control. In *Proceedings of the Fifth Annual Symposium on Human Interaction with Complex Systems* (pp. 87–91). Piscataway, NJ: IEEE.

Mosier, K. L., Palmer, E. A., & Degani, A. (1992, October). Electronic checklists: Implications for decision making. In Proceedings of the Human Factors and Ergonomics Society 36th Annual Meeting (pp. 7-11). Santa Monica, CA: Human Factors and Ergonomics Society.

Mosier, K. L., & Skitka, L. J. (1996). Human decision makers and automated decision aids: Made for each other? In R. Parasuraman & M. Mouloua (Eds.), Automation and Human Performance: Theory and Application (pp. 201-220). Mahwah, NJ: Erlbaum.

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

Moray, N., Inagaki, T.,&Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied, 6*, 44–58.

Moray, N., & Inagaki, T. (2000). Attention and complacency. Theoretical Issues in Ergonomics Science, 1(4), 354-365.

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. Ergonomics, 39(3), 429-460.

Nielsen, J. (1993). Usability engineering. San Francisco, CA: Morgan Kaufmann Publishers, Inc.

Norman, D.A. (1993). Chapter 3 - The Power of Representation. *Things that make us smart*. Reading, MA: Perseus Books.

Norman, D.A. (1988). The psychology of everyday things. New York: Harper & Row.

Oakley, I., Kim, Y., Lee, J., & Ryu, J. (2006) Determining the Feasibility of Forearm Mounted Vibrotactile Displays. Proceedings of the IEEE Conference on Virtual Reality, 74-81.

Olson, W.A. and Sarter, N.B. (2000). Automation Management Strategies: Pilot Preferences and Operational Experiences. *International Journal of Aviation Psychology, 10*(4), 327-341.

Olson, W.A. and Sarter, N.B. (2001). "Management-By-Consent in Human-Machine Systems: When and Why It Breaks Down". *Human Factors, 43*(2), 255-266.

Opperman, R. (1994). *Adaptive User Support*. Hillsdale, NJ: Erlbaum.

Palmer, M. T., Rogers, W. H., Press, H. N., Latorella, K. A., & Abbott, T. S. (1995). A crew-centered flight deck design philosophy for high-speed civil transportation (HSCT) aircraft (NASA Technical Memorandum No. 109171). Hampton, VA: NASA Langley Research Center.

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

Parasuraman, R., Bahri, T., Deaton, J., Morrison, J., & Barnes, M. (1992). Theory and design of adaptive automation in aviation systems. (Progress Report No. NAWCADWAR-92033-60). Warminster: Naval Air Warfare Center.

Parasuraman, R., Cosenzo, K. A., & Visser (2009). Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology, 21*, 270-297.

Parasuraman, R., Galster, S., Squire, P., Furukawa, H., & Miller, C. A. (2005). A flexible delegation interface enhances system performance in human supervision of multiple autonomous robots: Empirical studies with RoboFlag. *IEEE Transactions on Systems, Man&Cybernetics - Part A: Systems and Humans, 35*, 481–493.

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., & Miller, C. (2006). Delegation interfaces for human supervision of multiple unmanned vehicles: Theory, experiments, and practical applications. In N. Cooke, H. L. Pringle, H. K. Pedersen, & O. Connor (Eds.), *Human factors of remotely operated vehicles. Advances in human performance and cognitive engineering* (Vol. 7, pp. 251–266). Oxford, UK: Elsevier.

Parasuraman, R. Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced “complacency”. *International Journal of Aviation Psychology, 3*, 1-23.

Parasuraman, R., Mouloua, M., & Hilburn, B. (1999). Adaptive aiding and adaptive task allocation enhance human-machine interaction. In M. W. Scerbo & M. Mouloua (Eds.), *Automation technology and human performance: Current research and trends* (pp. 119–123). Mahwah, NJ: Erlbaum.

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

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

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

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

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

Perrow, C. (1986). Normal accidents: Living with high-risk technologies. New York: Basic Books.

Prinet, J., C., Terhune, A., & Sarter, N. B. (2012). Supporting dynamic re-planning in multiple UAV control: A comparison of 3 levels of automation. In *Proceedings of the Human Factors & Ergonomics Society 56th Annual Meeting* (pp. 423-427). Santa Monica, CA: Human Factors & Ergonomics Society.

Reason, J. (1997). Managing the risks of organizational accidents. Aldershot, UK: Ashgate.

Rosenthal, R. (1991). *Meta analytic procedures for social science research.* Newbury Park Cal.: Sage.

Rottger, S., Bali, K. & Manzey, D. (2009). Impact of automated decision aids on performance, operator behaviour and workload in a simulated supervisory control task. *Ergonomics*, *52* (5), 512–523.

Reeve, L. M., & Schmorrow, D. D. (2007). Augmented Cognition Foundations and Future Directions—Enabling "Anyone, Anytime, Anywhere" Applications. In Stephanidis, C. (Ed.): Universal Access in HCI, Part I, HCII 2007, LNCS 4554, 263–272.

Rouse, W. (1988). Adaptive interfaces for human/computer control. *Human Factors, 30*, 431–488.

Rouse, W. B., Geddes, N. D., & Curry, R. E. (1987) Architecture for interface: Outline of an approach to supporting operators of complex systems, *Human-Computer Interacion*, 3 (2), 87–122.

Sarter, N. B., & Schroeder, B. (2001). Supporting Decision Making and Action Selection under Time Pressure and Uncertainty: The Case of In-Flight Icing. *Human Factors, 43*(5), 573-583.

Sarter, N.B. and Woods, D.D. (1995). "How in the World Did We Ever Get Into That Mode?" Mode Error and Awareness in Supervisory Control. Human Factors, 37(1), 5-19.

Sarter, N.B. and Woods, D.D. (1994). Pilot Interaction with Cockpit Automation II: An Experimental Study of Pilots' Model and Awareness of the Flight Management System (FMS). International Journal of Aviation Psychology, 4(1), 1-28.

Sarter, N.B. and Woods, D.D. (1997). "Teamplay with A Powerful and Independent Agent": A Corpus of Operational Experiences and Automation Surprises on the Airbus A-320. Human Factors, 39(4), 553-569.

Sarter, N.B. and Woods, D.D. (2000). Teamplay with a Powerful and Independent Agent: A Full-Mission Simulation Study. Human Factors, 42(3), 390-402.

Sarter, N.,Woods, D., & Billings, C. E. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 1926–1943). New York: Wiley.

Scerbo, M.W. (2001). Adaptive automation. In W. Karwowski (Ed.), International encyclopedia of ergonomics and human factors (pp. 1077-1079). London: Taylor and Francis, Inc.

Scerbo, M. W. (2007). Adaptive automation. In R. Parasuraman & M. Rizzo (Eds.), *Neuroergonomics: The brain at work* (pp. 239–252). New York: Oxford University Press.

Sethumadhavan A. (2009). Effects of Automation Types on Air Traffic Controller Situation Awareness and Performance. In *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting* (pp. 1-5). Santa Monica, CA: Human Factors and Ergonomics Society.

Shneiderman, B. (1998). *Designing the user interface: Strategies for effective human-computer interaction (3rd ed.)*. Reading, MA: Addison-Wesley Publishing.

Sheridan, T. (1981). Understanding human error and aiding human diagnostic behavior in nuclear power plants. In J. Rasmussen & W. Rouse (eds.), Human Detection and Diagnosis of System Failures. New York: Plenum.

Sheridan, T. (1992). *Telerobotics, automation, and supervisory control.* Cambridge, MA: MIT Press.

Sheridan, T. B. (2002). *Humans and automation: Systems design and research issues.* New York: Wiley.

Sheridan, T. B., & Parasuraman, R. (2000). Human vs. automation in responding to failures: An expected-value analysis. *Human Factors, 42,* 403-407.

Sheridan, T., & Parasuraman, R. (2006). Human-automation interaction. *Reviews of Human Factors and Ergonomics, 1*, 89–129.

Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators.* (Technical Report). Cambridge, MA: Man Machine Systems Laboratory, MIT.

Sklar, A.E., & Sarter, N.B. (1999). Good vibrations: Tactile feedback in support of attention allocation and human-automation coordination in event driven domains. Human Factors, 41(4), 543-552.

Tan, H. Z., Gray, R., Young, J. J., & Traylor, R. (2003). A haptic back display for attentional and directional cueing. Haptics-e, 3, Article 1.

Tomlinson, Z. A. (2009). Influence of spatial abilities on primary and secondary space telerobotics operator performance. M.S. thesis. Massachusetts Institute of Technology.

van Erp, J. B. F. (2001). Tactile navigation display. In S. Brewster & R. Murray-Smith Eds.), Haptic HCI 2000 (pp. 165-173). Berlin: Springer-Verlag.

van Erp, J. B. F. (2005). Presenting directions with a vibrotactile torso display. Ergonomics, 48. 302-313.

Wickens, C.D. (2008). Multiple resources and mental workload. Human Factors, 50(3), 449-455.

Wickens, C.D., Lee, J., Liu, Y., & Gordon Becker, S. (2004). An Introduction to Human Factors Engineering (2nd edition). Upper Saddle River, NJ: Prentice Hall.

Wickens, C. (1994). Designing for situation awareness and trust in automation. In Proceedings of the IFAC Conference. Baden-Baden, Germany, pp. 174-179.

Wickens, C.D. (2008). Multiple resources and mental workload. *Human Factors Golden Anniversary Special Issue, 3,* 449-455

Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science 8,* 201-212.

Wickens, C.D, & Carswell, C.M. (1995). The proximity compatibility principle: Its psychological foundation and its relevance to display design. *Human Factors, 37*(3), 473-494.

Wickens, C.D., Gordon, S., and Liu, Y. (1998). Introduction to Human Factors Engineering. New York: Addison-Wesley-Longman.

Wickens, C. D., Hollands, J. G. (2000). Engineering Psychology and Human Performance. Upper Saddle River, NJ: prentice-Hall Inc.

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 54th Annual Meeting (pp. 389-393). Santa Monica, CA: Human Factors and Ergonomics Society.

Wickens, C. D., Mavor, A., Parasuraman, R., & McGee, J. (1998). *The future of air traffic control: Human operators and automation.* Washington DC: National Academy Press

Wiener, E.L. (1988). Cockpit automation. In E.L. Wiener & D.C. Nagel (Eds.), *Human factors in aviation* (pp. 433-461). San Diego: Academic Press.

Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics, 23,* 995-1011.

Wright, P. (1974). The harassed decision maker: Time pressures, distractions, and the use of evidence. Journal of Applied Psychology, 59, 555-561.

Woods, D. (1996). Decomposing automation: Apparent simplicity, real complexity. In R. Parasuraman, M. Mouloua, & R. Molloy (Eds.), Automaton and human performance: Theory and applications (pp. 3–17). Mahwah, NJ: Erlbaum.

Wilson, G., & Russell, C. A. (2007). Performance enhancement in a UAV task using psychophysiologically determined adaptive aiding. *Human Factors, 49,* 1005–1018.

Wright, M.C. & Kaber, D.B. (2005). Effects of automation of information-processing functions on teamwork. *Human Factors*, *47* (1), 50–66.

Yeh, M., Merlo, J.L., Wickens, C.D., & Brandenburg, D.L. (2003). Head up vs. head down: The costs of imprecision, unreliability, and visual clutter on cue effectiveness for display signaling. *Human Factors, 45*(3), 390-407.

Yeh, M., & Wickens, C. D. (2001). Display signaling in augmented reality: Effects of cue reliability and image realism on attention allocation and trust calibration. Human Factors, 43, 355–365.