Spring 2024

Visual Attention in Remote Vehicle Supervision: Examining the Effects of Mental Models and Information Bandwidth

Michael Stanley Politowicz
Old Dominion University, michaelpolitowicz@gmail.com

VISUAL ATTENTION IN REMOTE VEHICLE SUPERVISION: EXAMINING THE EFFECTS OF MENTAL MODELS AND INFORMATION BANDWIDTH

by

Michael Stanley Politowicz
B.S.E. May 2011, University of Michigan

A Thesis Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

PSYCHOLOGY

OLD DOMINION UNIVERSITY
May 2024

Approved by:

Yusuke Yamani (Director)

Xiao Yang (Member)

Michelle L. Kelley (Member)

ABSTRACT

VISUAL ATTENTION IN REMOTE VEHICLE SUPERVISION:
EXAMINING THE EFFECTS OF MENTAL MODELS AND INFORMATION BANDWIDTH

Michael Stanley Politowicz
Old Dominion University, 2024
Director: Dr. Yusuke Yamani

Advances in automation and aviation technologies have been catalysts for the emerging market of Advanced Air Mobility (AAM), an ecosystem of novel aircraft concepts including package delivery drones and passenger carrying air-taxis. Future aircraft operators in this environment will be tasked with remotely supervising multiple highly automated aircraft on a visual interface while receiving less training than traditional pilots. More research should explore how an operator's potentially limited understanding of an automated system affects visual performance and interactions between human operators and AAM technologies. This study examined the influence of mental models of an autopilot system on visual attention allocation for participants managing multiple vehicles in a low-fidelity AAM simulation environment. Fifty-five participants completed a series of multi-aircraft control scenarios after reading training slides with or without explicit information on the underlying functionality of an autopilot system (Advanced Mental Model or Basic Mental Model groups, respectively) with their eye movements recorded. The results indicated that the Advanced Mental Model group allocated significantly more visual attention to a supplemental data display than the Basic Mental Model group. Surprisingly, participants allocated more visual attention to the supplemental data display with low than high information bandwidth, which was opposite of the predicted effect. The results also indicated a significant interaction between expectancy and value parameters in the

SEEV model, providing additional evidence in support of this theoretical debate. In practice, results from this study show that refining mental models through a simple training program could be an effective approach to alter AAM operators’ visual scan behaviors.

This work is dedicated to Bri, Trousers, Loafers,
my family, friends, and colleagues.

ACKNOWLEDGEMENTS

I would first like to acknowledge and thank my advisor and mentor, Dr. Yusuke Yamani. His guidance, knowledge, and encouragement have pushed me to exceed even my own expectations as a researcher, and I extend my sincere appreciation for taking a chance on an aerospace engineer with no background in psychology. I am looking forward to what lies ahead. I would also like to acknowledge and thank my NASA colleague and collaborator, Dr. Eric Chancey, whose enthusiasm for human factors psychology ignited my own passion and led me to embark on this academic journey at ODU. I also extend a special thanks to the members of my committee, Dr. Xiao Yang and Dr. Michelle Kelley, for their time, effort, and thoughtful feedback. Their expertise has helped shape this thesis document.

Additionally, I greatly appreciate the support of my peers at ODU, including Tetsuya Sato, Jeffrey Glassman, Dr. James Unverricht, Dr. Sarah Yahoodik, Austin Jackson, Soyeon Lee, and Sam Petkac, for their enthusiasm and feedback leading up to my defense of this thesis. I also extend my appreciation to my colleagues at the NASA Langley Research Center, especially those in the Crew Systems and Aviation Operations Branch (CSAOB) and Transformational Tools and Technologies (TTT) project, for their support, encouragement, and inspiration.

Finally, I would like to thank my parents for always encouraging me to take risks in pursuit of my passions. And most importantly, I would like to acknowledge my partner, Bri Weadock, for her endless support through many late nights and early mornings. Work, school, and life balance has been a formidable challenge made easier by her unwavering care, and I am forever grateful. And of course, to our wonderful pets, Trousers and Loafers, who bring me joy every day and keep me company during long nights of writing.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Advances in automation technology in modern aviation systems are expected to fundamentally change how humans interact with automated systems, which are becoming increasingly more complex and prevalent (Endsley, 2017). Such advances in automation and other technologies (e.g., electric propulsion) have been catalysts for the emerging market of Advanced Air Mobility (AAM), which includes novel aircraft concepts such as package delivery drones and passenger carrying air-taxis (National Academies of Sciences, Engineering, and Medicine, 2020). Future aircraft operators (i.e., pilots) in the AAM environment will be tasked to remotely manage multiple aircraft while receiving less training than traditional pilots (Aubuchon et al., 2022; Patterson et al., 2021), bringing in some advantages associated with the increased prevalence, complexity, and authority of automation (de Winter, 2019). Yet, this new role of a human operator, with less experience and knowledge than traditional pilots, also presents several challenges. Remote operations introduce the possibility of degraded mental states of situation for the operator compared to onboard pilots (see Mutzenich et al., 2021). Additionally, monitoring multiple aircraft simultaneously may be restricted by attentional limitations of the operator (see Cavanagh & Alvarez, 2005). Reduced training and thus a lack of expertise may negatively affect visual scanning behavior (Brams et al., 2019), which presumably reduces the quality of their perception and comprehension of the automated system's state. Lastly, and importantly, higher degrees of automation may decrease operator performance in responding to automation failures (Onnasch et al., 2014).

Human-in-the-loop simulation study is necessary to explore psychological mechanisms that are responsible for supporting seamless interactions between the human operator and AAM technologies. For example, as the aircraft operator role evolves from an active to passive role with reduced training, the operator must demonstrate a minimum level of understanding about a vehicle's automated systems so that the operator can effectively and safely operate the vehicle (see Goodrich & Boer, 2003). Limited operator knowledge may lead to poor visual scanning behavior (Brams et al., 2019; Sarter et al., 2007). In the AAM environment, where operator training and skills are expected to be less comprehensive than those of traditional pilots, researchers will need to understand and address the potential impacts of operator mental models of air vehicles on visual attention allocation. The objective of the current study was to determine how a non-expert's mental model of an automated system affects visual attention allocation using a testbed for multitasking that reflects a remote aircraft operator's task load.

MENTAL MODELS OF AUTOMATION

Mental models are operators' internal and functional representations of automated systems. *Mental models* can be defined as "the mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states" (Rouse & Morris, 1986, p. 351). Mental models can provide enough detail for an operator to describe the purpose, explain the functions, and predict the behaviors of an automated system. However, a mental model is merely a simplified, yet functional, representation of a real system implying that finer details are lost in the mapping of the real system to the model (Moray, 1999). Hence, these models are not complete but approximations of the target system at a level that is sufficient for guiding operator behaviors (Moray, 1999; Norman, 1983; Norman, 1986). There is inherent value in employing an approximate model, including the ability to guide system behavior efficiently and appropriately without requiring extensive knowledge of the system (Norman, 1986).

Because mental models often represent only a subset of the information about a target system, the specific nature of the content within a model is important yet variable depending on the individual's goal and the situation. Furthermore, as automated systems become more complex and increasingly difficult to understand, research is necessary to identify what information about an automated system an operator must retain to safely and effectively use the system. To address this question, it is first important to understand the attributes of an automated system. *Automation* can be defined as "a device or system that accomplishes (partially or fully) a function that was previously, or conceivably could be, carried out (partially or fully) by a human operator" (Parasuraman et al., 2000, p. 287). However, this definition does not define any characteristics of automation that help frame a user's understanding of the system. Lee and Moray (1992; Lee & See, 2004), on the other hand, proposed an attributional abstraction of automation comprised of three central attributes of an automated system: *purpose*, *process*, and *performance*. *Purpose* refers to the designer's intent for the automation (i.e., *why* the automation was developed); *process* refers to the underlying algorithms of the automation, which can include physical form (i.e., *how* the automation operates); and *performance* refers to the current and past operation of the automation (i.e., *what* the automation does). These three attributes of automation closely align with the three functions included in the definition of mental models proposed by Rouse and Morris (1986) that was previously introduced. For the current study, these automation attributes provide a convenient basis for categorizing information presented to the human to support the formation of mental models.

For future aircraft operators, access to process-based information is expected to decrease due to limited understandability of future automation and reduced training requirements. Limiting access to process-based information can negatively affect operator performance (Kieras & Bovair, 1984). Furthermore, an inaccurate understanding of process-based information can also negatively impact operator performance (Sarter et al., 2007). Yet, additional research is required to understand the potential impact of reduced availability of information on operators' mental models and performance in the context of AAM. The current study specifically addressed operator access to process-based information.

It is necessary for researchers to understand characteristics of mental models to control and predict performance of an operator interacting with a target system. Norman (1983; 1986) provides another framework for describing mental models using three distinct models: the design model, the system image, and the user's model. The *design model* is the conceptual model of the system that is held by the designer of that system. Conversely, the *user's model* is the conceptual model of the system formed by the user (i.e., the mental model). Ideally, the goal of the designer should be to align the user's model with the design model by constructing an effective system image. The *system image* is the physical representation of the system, which includes everything that the user interacts with (e.g., user interfaces, documentation, training materials). Note that the user's model is not directly formed from the design model, but rather, the user's model evolves with exposure to the system image. For users with limited exposure to a system's user interfaces (i.e., non-experts), the system image is predominantly represented by documentation and training materials. Thus, training materials can be an effective tool for adjusting a non-expert user's mental model.

Previous studies have successfully altered a user's model of an unfamiliar system using documentation and training materials (Weis & Wiese, 2022; Politowicz et al., 2022). For example, with an unfamiliar smartphone app, Weis and Wiese (2022) found that a single paragraph describing the task-specific capabilities of the app was enough to substantially adjust participants' mental models and subsequent behavior. Additionally, Politowicz and colleagues (2022) found differences in participants' mental models of an unfamiliar autopilot after a brief training manipulation.

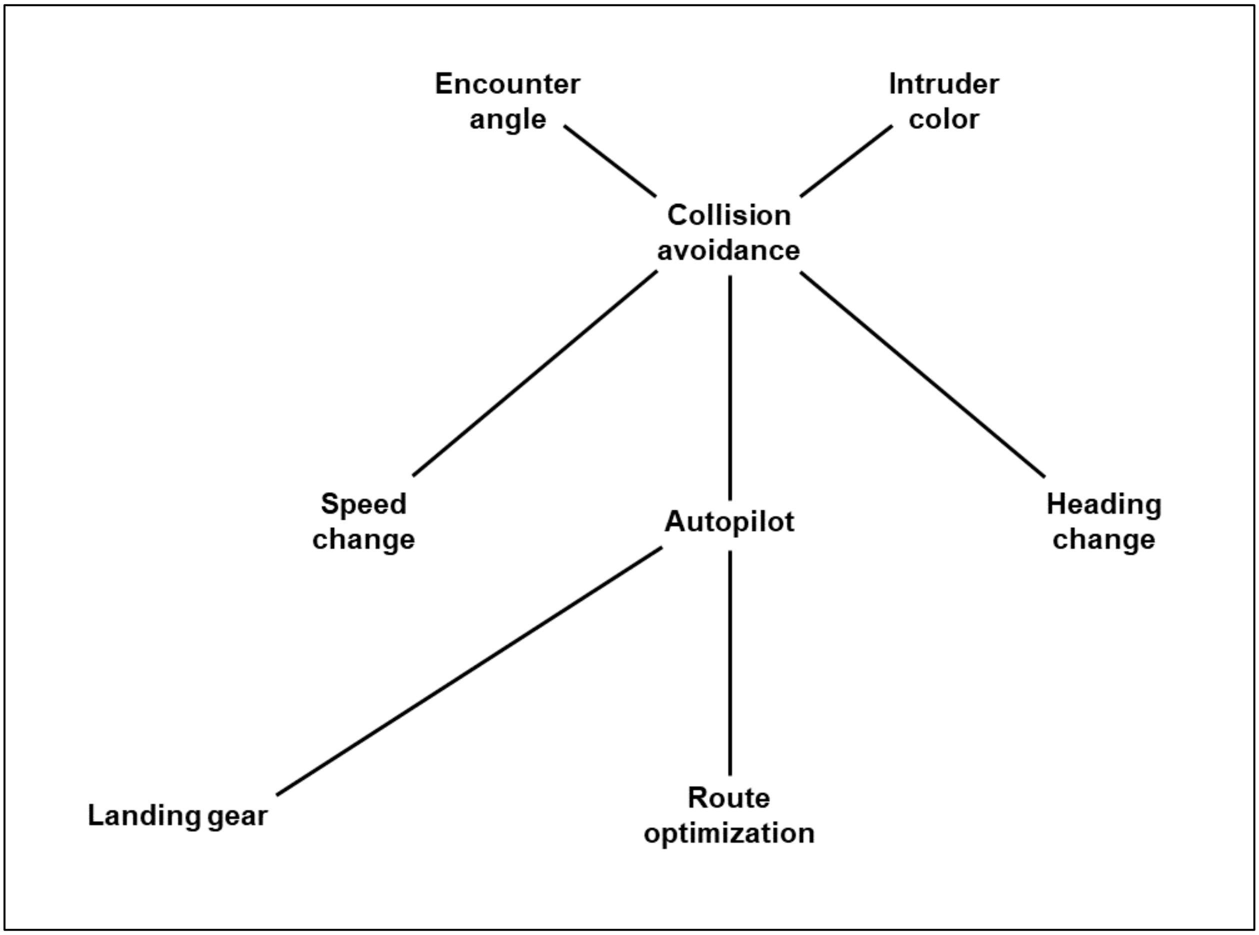
To this point, mental models and how they can be altered have been defined, but how mental models can be represented and elicited has not yet been discussed. There are many methods available for elicitation and representation of mental models (Rouse & Morris, 1986; see Hoffman et al., 2018, p. 10-12, for list of elicitation methods). The current study used Pathfinder network analysis (Schvaneveldt et al., 1989) for representing and analyzing mental models, because Pathfinder network analysis was found useful and independently predictive of performance compared to three other available methods (Rowe & Cooke, 1995). Additionally, using Pathfinder network analysis, Politowicz and colleagues (2022) were able to find differences in mental models for an autopilot system similar to the system used in the current study.

PATHFINDER NETWORKS

Pathfinder network analysis is an algorithm that generates networks to visualize and analyze the organization and relationships of concepts based on proximity data, which are derived from pairwise relatedness ratings obtained from participants (Schvaneveldt et al., 1989; see Figure 1). This method provides the mechanisms for generating networks regardless of the network content. In other words, the method serves as a framework for analyzing data rather than a test of specific content. The network content is comprised of a list of concepts (i.e., nodes in the network) that must be defined prior to collecting pairwise relatedness ratings from participants. Thus, Pathfinder is a versatile method that has been used to evaluate mental models across multiple focus areas, including team mental models (Lim & Klein, 2006), system troubleshooting (Rowe & Cooke, 1995), and video game expertise (Furlough & Gillan, 2018). However, Pathfinder networks have not previously been applied to studying mental models of automated systems using the theoretical framework proposed in the current study (i.e., purpose-based, process-based, and performance-based information). Although, it is important to note that previous mental model studies employing other theoretical approaches have found the Pathfinder method to be predictive of expertise (Furlough & Gillan, 2018; Schvaneveldt et al., 1985).

Figure 1

Example Pathfinder Network from Politowicz et al. (2022)



Note. The concepts (or nodes) included in this example Pathfinder network are not the concepts that were used in the current study. From "Pathfinder Networks for Measuring Operator Mental Model Structure with a Simple Autopilot System," by M. S. Politowicz, T. Sato, E. T. Chancey, and Y. Yamani, 2022, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), pp. 883–887. In the public domain.

A network consists of a set of *nodes*, which correspond to predefined keywords or concepts. The nodes in a network are connected by weighted *links*, which represent the distance between (or relatedness of) two nodes. For example, in Figure 1, the concepts *heading change* and *collision avoidance* (represented as nodes) are depicted as closely related because there is a link directly connecting the two concepts. Although the network visualization does not depict a weight value for this link, these weights are fundamental for determining the overall network structure. Pathfinder is advantageous because it reduces the full set of pairwise relatedness ratings to only the "strongest" connections. This reduction is accomplished by removing the "weak" links between nodes if there is a stronger connection available through some combination of other links. Thus, the distance between two nodes is defined by the weight of the shortest path between those nodes (Dearholt & Schvaneveldt, 1990). For example, in Figure 1, although the data used to generate this Pathfinder network contained a relatedness rating for *autopilot* and *heading change*, the network does not include a direct link between these concepts because the connection through *collision avoidance* was stronger than the direct link (i.e., the relatedness rating for *autopilot* and *heading change*).

Two parameters are required as inputs to the Pathfinder analysis: the Minkowski *r*-metric (also referred to as the order of the Minkowski distance) and the *q* parameter. The *r*-metric determines how weights between nodes are computed, and *q* is the maximum number of links allowed in a path between two nodes. Networks become sparser (i.e., have fewer links) as either parameter increases. The most commonly used values, which together yield the fewest number of links, are $r = \infty$ and $q = (n - 1)$, where *n* is the number of nodes in the network (Dearholt & Schvaneveldt, 1990). To calculate the weight of each possible path between two nodes, the weight, *W*, of path, *P*, is defined as:

$$W(P) = \left(\sum_{i=1}^{k} w_i^r\right)^{\frac{1}{r}} \tag{1}$$

where w_i is an individual link weight along the path of k links. For $r = \infty$, the path weight is simply the maximum weight of any individual link along the path.

The Pathfinder network produced from this method provides a useful graphical representation of the mental model, but measures of interest can also be derived from the network and the relatedness ratings. *Network similarity*, *coherence*, and *internal consistency* are common Pathfinder metrics that are useful in the context of mental model research and were included in this study. *Network similarity* is a metric for comparing two networks by evaluating common links across the networks, which accounts for both shared and unshared links. *Coherence* measures the reliability of an individual's relatedness ratings, and *internal consistency* is a binary value based on a threshold value of coherence. Further details and calculations for these metrics are provided in the description of dependent measures in a subsequent section.

MODEL OF VISUAL ATTENTION ALLOCATION

Vision is a primary information-processing channel that supports operator performance in aviation (e.g., Sarter & Woods, 1995; Wickens et al., 2008; Ziv, 2016). Computational models have been developed to characterize and predict visual scanning patterns and performance for operators in the domain of aviation human factors. For models of supervisory control, the eye is conceptualized as a single-server queue, and movement of the eyes sequentially serves this queue (Senders, 1964; see Salvucci & Taatgen, 2008). In other words, information access in the visual channel is limited by the constraint that an operator's eyes can only fixate in a single location at a time. These sequential fixations are often used as a measure of attention allocation (e.g., Karpinsky et al., 2018; Salvucci & Taatgen, 2008; Sato et al., 2020). Though the locations of fixations and attention can become decoupled (e.g., Posner et al., 1980), models using eye movement behavior to reflect visual attention allocation have been effective in explaining operator monitoring behaviors for supervisory control tasks (e.g., Horrey et al., 2006).

The *salience-effort-expectancy-value* (or *SEEV*) model for supervisory control provides a computational method for predicting gazes towards predefined areas of interest (AOIs) based on the four parameters of the model (i.e., the primary forces that move visual attention; Wickens, 2015; Wickens et al., 2022). An attentional weight for each AOI, A_i, can be calculated using the following formula:

$$A_i = sS_i - efEF_i + (exEX_i \times vV_i) \quad (2)$$

where *i* represents the index of each AOI, the uppercase letters represent the characteristics of each AOI (typically set to ordinal values from low to high across AOIs; see Wickens, 2015, for computed example), and the lowercase letters represent weighting values based on the operator's attentional strategy (typically all set to 1 by default; Wickens et al., 2022). *S* and *s* refer to salience; *EF* and *ef* refer to effort; *EX* and *ex* refer to expectancy; and *V* and *v* refer to value. This equation can be used to predict and compare the relative distribution of attention across AOIs.

In the SEEV model, *salience* and *effort* are considered bottom-up properties because these are associated with physical characteristics of the display interface. Conversely, *expectancy* and *value* are considered top-down properties as these are dependent upon the mental model of the operator. *Salience* refers to the extent to which an AOI stands out from other AOIs based on size, color, intensity, or contrast (Wickens et al., 2013). AOIs with higher salience tend to draw more attention, as indicated by the positive contribution of sS_i in Equation 2. *Effort* refers to the cost of moving attention from one AOI to another, which corresponds to the degree of spatial separation (i.e., physical distance) between AOIs (Wickens et al., 2013). An AOI that requires more effort to access (i.e., eye, head, or body movement; information access cost) tends to draw less attention, as indicated by the negative contribution of $efEF_i$ in Equation 2. For the current study, salience and effort properties of the display remained fixed to enable a specific focus on the influence of the two top-down sources of attention guidance: *expectancy* and *value*.

Expectancy corresponds to the information bandwidth, or rate of change, of information within an AOI (Wickens et al., 2013). Although information bandwidth is a physical characteristic of the display, expectancy is considered a top-down property because it specifically refers to the operator's expectation about the rate of change within an AOI, which becomes calibrated to actual information bandwidth with experience (see Underwood et al., 2002). For example, during the descent phase of flight, an expert aircraft pilot will likely expect the altimeter gauge to have high information bandwidth because, during descent, the altitude changes rapidly. The altimeter should therefore draw more attention during descent as a result of this expectation. Hence, expectancy will guide attention in correlation with the information bandwidth of an AOI, such that higher expected bandwidth tends to draw more attention (Senders, 1964).

The other top-down property of the SEEV model is *value*, which refers to the importance of information within an AOI. *Value* is defined by the relevance of an AOI to a specific task weighted by the priority of that task relative to other tasks (Wickens et al., 2013). Horrey and colleagues (2006) examined the effect of task priority in a simulated driving experiment by instructing participants to change task priorities between trials. Their results were consistent with SEEV model predictions for task priority. On the other hand, task relevance of an AOI is based on the operator's understanding of the information displayed within that AOI, which in some contexts could be influenced by the operator's mental models of the systems involved in the execution of the task. More broadly, expertise of an operator has been shown to influence both value and expectancy (i.e., the top-down properties of the SEEV model). For example, using the SEEV model, Koh and colleagues (2011) found that experts allocated attention more effectively than novices.

Based on Equation 2 proposed by Wickens and colleagues (2022), the value and expectancy properties of the SEEV model are assumed to interact with one another. One interpretation of this interaction is that an AOI with high importance (i.e., high value) that never changes (i.e., no expectancy) will not draw any attention from the operator. Conversely, an AOI with no importance (i.e., no value) that changes frequently (i.e., high expectancy) also will not draw any attention from the operator. In simpler terms, if an operator does not care about the information being presented, then the amount of time spent looking at that information will not be influenced by how often the display updates. This equation and interaction are consistent with the original SEEV model proposed by Wickens and colleagues (2001), as well as the model used by Wickens and colleagues (2008). However, the study conducted by Wickens and colleagues (2008) was a validation study that found the data fit the SEEV model better with expectancy and value as additive effects rather than interacting effects. Following this finding, Wickens and McCarley (2008) and Wickens (2015, 2021) presented the SEEV model equation with expectancy and value as additive effects rather than interacting effects. Wickens (2015) justified the additive model with two specific reasons. First, the validation study of the SEEV model with experienced pilots (Wickens et al., 2008) found that the additive model fit the data better than the multiplicative model. Second, for a real supervision task of a complex system, having both expectancy and value terms go to zero when just one of the terms is zero is not a rational strategy. Yet, despite this justification, the latest model proposed by Wickens and colleagues (2022) has reverted to the original form of the model (i.e., Equation 2). The current study provides empirical data to address this discrepancy in the literature by directly testing the relationship between expectancy and value.

STUDY PURPOSE AND RESEARCH QUESTIONS

The current study examined the influence of mental models of an autopilot system on visual attention allocation for participants acting as remote aircraft operators (i.e., pilots) managing multiple vehicles in a low-fidelity flight simulation environment. Previous studies have shown that participants with substantial mental model differences (i.e., novices vs. experts) allocate visual attention differently (Brams et al., 2019). Additionally, studies have shown that mental models can be manipulated easily, and these differences lead to altered behavior (Weis & Wiese, 2022). Yet whether and how a simple manipulation of the mental model of an automated system (i.e., purpose-based, process-based, and performance-based information) affects visual attention allocation has not been explored. The literature indicates that research addressing the effect of top-down manipulations on visual attention allocation has focused on expertise and task-based properties (e.g., task priority) rather than system properties (e.g., Horrey et al., 2006). This study examined whether a simple manipulation of the mental model of an automated system leads to predicted differences in visual attention allocation based on the SEEV model. Furthermore, this study investigated the relationship between expectancy (i.e., information bandwidth) and value (i.e., mental model) parameters of the SEEV model.

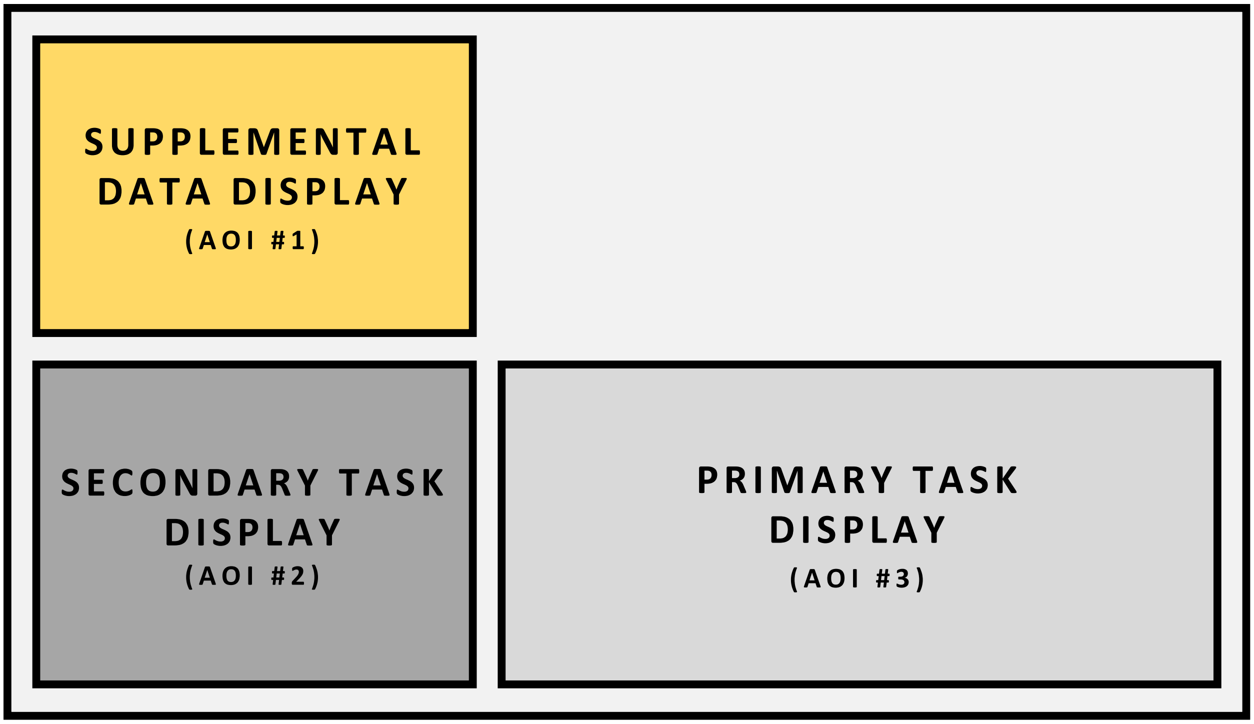
The current study used a testbed for multitasking that resembles a remote aircraft operator's task load. During a series of dual-task scenarios, participants were responsible for a resource management task (primary task) and a system monitoring task (secondary task). The resource management task included two separate aircraft that each employed automation (i.e., autopilot). Supplemental autopilot status data was presented in a supplemental data display, and all participants had access to this display. However, only half of the participants were provided with additional process-based information about the autopilot (advanced mental model) that provided context for the value of the information displayed in the supplemental data display. The other half of the participants did not receive this additional information (basic mental model). The configuration of the displays is depicted in Figure 2. Additionally, the information bandwidth for the supplemental data display was altered between low and high settings across scenarios as a within-subjects manipulation. The following hypotheses were proposed:

> H1: Operators with advanced mental models will allocate more visual attention to a supplemental data display than operators with basic mental models.
>
> H2: Operators will allocate more visual attention to a supplemental data display in the high information bandwidth condition than the low information bandwidth condition.
>
> H3: The effect of the mental model manipulation will be larger in the high than low information bandwidth conditions, only for those with advanced mental models but not for those with basic mental models.

Additionally, the current study included exploratory analyses targeting the effects of mental models on task performance as well as the effectiveness of Pathfinder network analysis for detecting differences in mental models in the context of AAM.

Figure 2

Display Configuration of the Experimental Multitasking Testbed



Note. This diagram represents the relative positions of the displays on the user interface in the current study.

CHAPTER II

METHOD

PARTICIPANTS

Fifty-five participants were recruited from the community of students at Old Dominion University. Twenty-seven participants were randomly assigned to the Basic Mental Model group (M = 19.74 years, SD = 4.95, 20 female), and 28 participants were randomly assigned to the Advanced Mental Model group (M = 20.25 years, SD = 5.26, 21 female). Participants were screened for normal color perception and normal or corrected-to-normal visual acuity. Participants reported video game use of 4.95 hours per week on average, and a two-tailed, independent-samples t-test indicated that there was no significant difference in video game use between the Basic (M = 6.07, SD = 9.73) and Advanced (M = 3.88, SD = 6.12) Mental Model groups, independent-samples $t(53) = 1.01$, $p = .318$, $d = 0.27$.

One additional person participated, but their data are excluded because the eye tracking data indicated that they were not viewing the screen for a majority of the time during each trial. No additional exclusion criteria were used. This research was approved by the Institutional Review Board (IRB) Committee of Old Dominion University (approval #1482888-4).

DESIGN

The experiment employed a 2 x 2 split-plot factorial design. The between-subjects factor was Mental Model, which differed in autopilot system knowledge exposure during training. The within-subjects factor was Supplemental Data Display Bandwidth (Low vs. High), which differed in the frequency of information display changes.

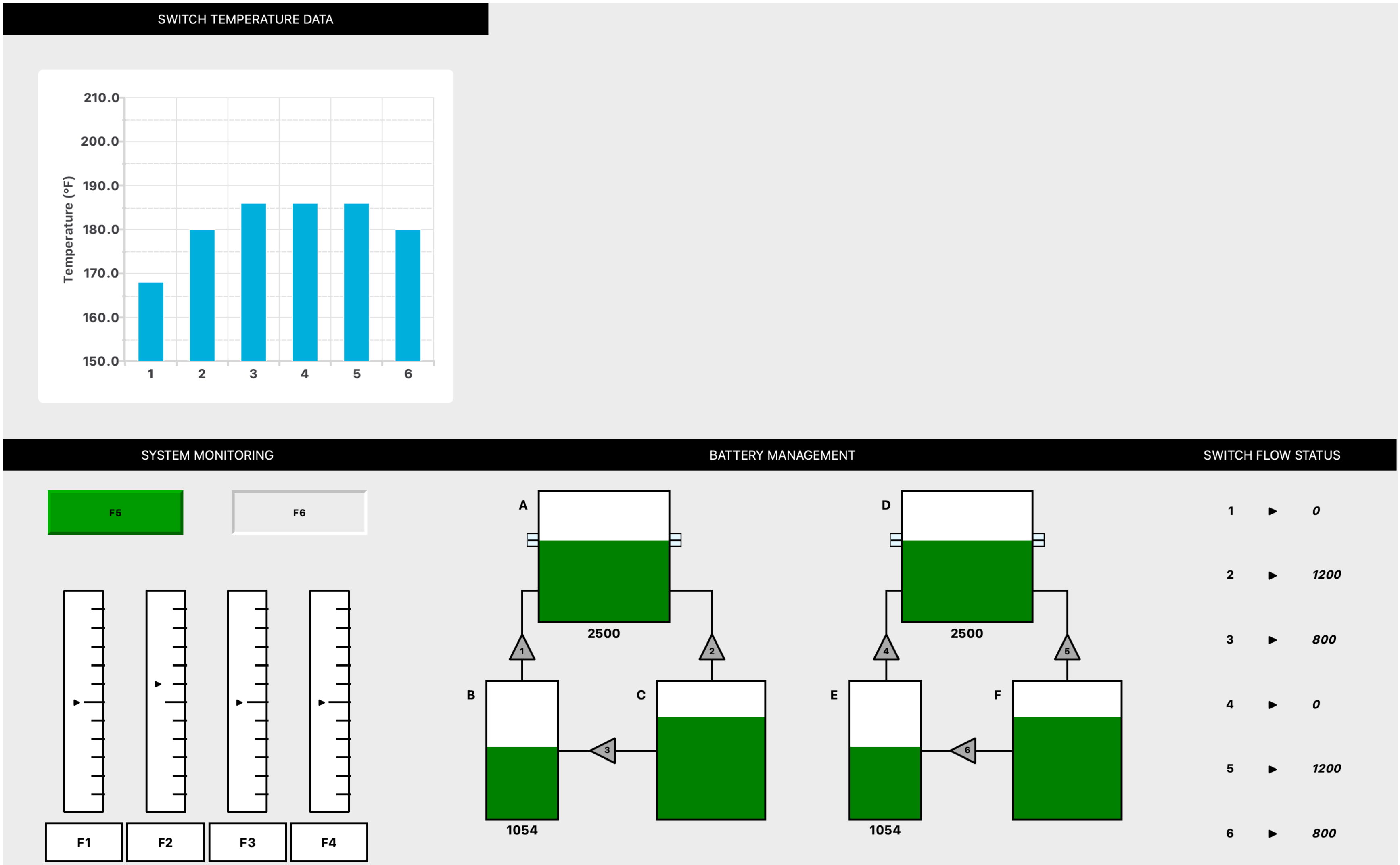
APPARATUS AND MATERIALS

OpenMATB

OpenMATB (Cegarra et al., 2020) is an open-source replication of the original Multi-Attribute Task Battery (MATB) software developed by the National Aeronautics and Space Administration (NASA; Comstock & Arnegard, 1992). The OpenMATB is a computer-based simulation environment representative of a low-fidelity aircraft pilot interface. The software provides a multitask environment, and the tasks that were employed in this study are the Resource Management Task and the System Monitoring Task (Figure 3). The Resource Management Task requires the participant to manage specific levels of a resource across multiple containers. This task was modified for the current study to represent multiple vehicles rather than a single vehicle, and the behavior of the automation used in this task was modified to accommodate the study design. For this task, the participant has to monitor the resource levels and determine which resource transfer switch is malfunctioning (i.e., needs to be clicked) when the automation fails. The System Monitoring Task requires the participant to monitor several indicators and click a corresponding button if one changes to an off-nominal state. Both tasks are highly configurable by the researcher.

Figure 3

Participant User Interface for the OpenMATB Software



Note. This example user interface includes the Resource Management Task (bottom center), System Monitoring Task (bottom left), and Supplemental Data Display (top left).

This study also employed the novel Supplemental Data Display, which is connected to the Resource Management Task and shows corresponding graph-based information. This display was developed specifically for the current study. Information in this display provides supplemental information related to the state of the Resource Management automation (i.e., autopilot). Specifically, the Supplemental Data Display is configured to indicate when the autopilot and the corresponding resource transfer switches are malfunctioning. The display refresh rate for the Supplemental Data Display was set to a high frequency (5 Hz) for the High information bandwidth condition and a low frequency (1 Hz) for the Low information bandwidth condition.

Eye Tracker

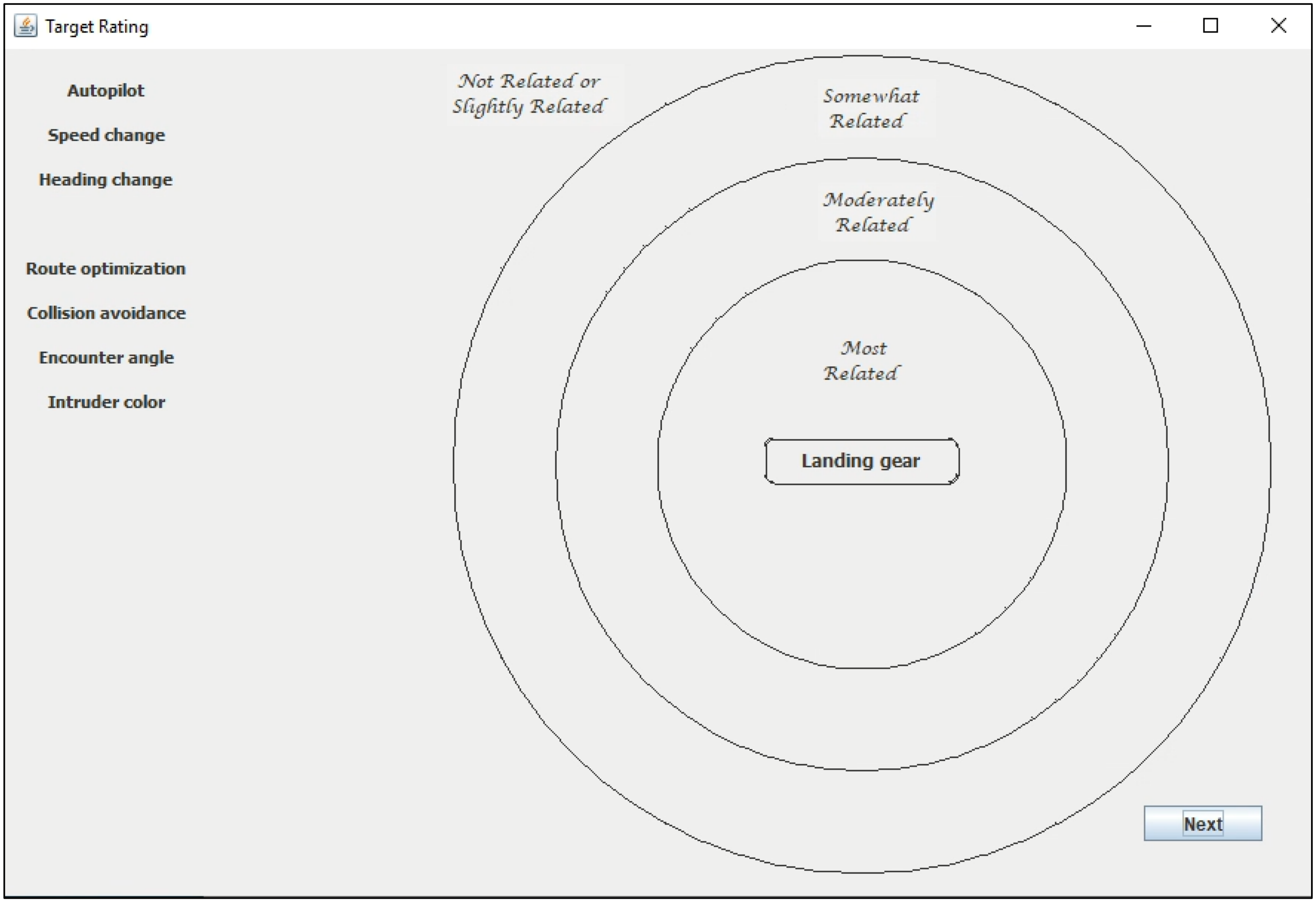
A Tobii Pro Nano screen-based eye tracker was used to record the participant's eye movement data with a sampling rate of 60 Hz.

Target Rating Tool

To elicit participants' mental models, a user interface for collecting pairwise relatedness ratings with the Target Rating method was used (JTarget; Interlink, 2017b; Figure 4). The Target Rating method was developed by Tossell and colleagues (2010) and was found to be an efficient method for collecting comparison ratings. Politowicz and colleagues (2022), for example, were able to detect differences in participants' mental models of a simplified autopilot system based on a simple exposure training intervention using the Target Rating method. The Target Rating method uses concentric circles resembling a target to represent levels of relatedness to the concept in the center circle (see Figure 4). Participants move all other concepts from the set into the circles that correspond to degree of relatedness with the primary concept. When the participant completes ratings for a concept, the target view resets, and a new primary concept appears in the center circle. This repeats until all concepts have cycled through the center circle position. The target rating tool randomizes the order of target keywords presented to the participant.

Figure 4

Target Rating Method User Interface (JTarget Software) used in Politowicz et al. (2022)



Note. The concepts (or keywords) included in this example user interface are not the concepts that were used in this study. From "Pathfinder Networks for Measuring Operator Mental Model Structure with a Simple Autopilot System," by M. S. Politowicz, T. Sato, E. T. Chancey, and Y. Yamani, 2022, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), pp. 883–887. In the public domain.

Training Materials

Both groups were introduced to the concept of Advanced Air Mobility (AAM), the details of the experimental task, and the high-level functionality of an autopilot. The autopilot discussed in the training was fictitious and simplified (relative to a conventional aircraft autopilot), but the autopilot description aligned with the behavior of the vehicle automation in the Resource Management Task. The only difference in training content between the two mental model groups was the inclusion of additional low-level autopilot functionality (i.e., process-based) information for the Advanced Mental Model group.

For the Basic Mental Model group, general details about AAM were introduced to provide context for the tasks. Next, the tasks were described with sufficient detail for the participants to effectively complete the study. Participants were introduced to the behavior and functionality of the autopilot at a high level. Participants were also informed that the battery management feature of the autopilot is not perfectly reliable. Scores indicating performance were not displayed for participants during the scenarios. Participants were instructed to complete the task as accurately and quickly as possible. In the final section of the training, the Basic group was provided with additional details about AAM that were not relevant to the experimental tasks, such as the expected economic impact of AAM.

For the Advanced Mental Model group, the training was identical to the Basic group except for four slides in the final section. Rather than receiving additional information about AAM in the final section, the Advanced group was introduced to low-level functionality of the autopilot (i.e., process-based information) that provided context for the utility of information displayed in the Supplemental Data Display. Specifically, participants in both groups were told that the Supplemental Data Display shows the temperature of each battery switch on the aircraft.

Participants in the Advanced group were then instructed that the autopilot's battery management algorithm (for distributing power to each battery) becomes confused and fails when it cannot detect the current status of a switch due to overheating (i.e., temperature rises above a certain threshold value). From this information, the Advanced group participants can infer which switch is malfunctioning when the Supplemental Data Display shows a temperature value above the threshold. Importantly, the implementation of the supplemental data did not align with presumed existing mental models of the participants to reduce the likelihood of this information being understood without explicit training. Additionally, because eye gaze position was used as the approximate measure for visual attention allocation, the supplemental data was presented in a manner that generally requires sustained focal vision for perception (e.g., reading text) rather than a form that could be more easily perceived with ambient vision or brief eye glances. Furthermore, the Advanced group training did not directly describe the application of this supplemental information to the experimental task to avoid introducing differences in task instructions between groups (i.e., mental model of the task). Rather, the Advanced Mental Model training required participants to make inferences about the task based on their mental model of the autopilot.

The training was presented to participants as slides on the computer. The training slides used for both groups are included in Appendix A. Both mental model groups were shown an equal number of slides (79 total), but four of the slides were different between groups. To validate the effectiveness of the training intervention, participants were required to correctly answer twelve training assessment multiple-choice questions (specific to each mental model group) immediately following the training (Appendix B). If the participant answered any question incorrectly, they were asked to review the training materials again and then retake the same assessment. The participant repeated this process until all questions were answered correctly. After the second failed attempt, participants were given access to the training slides and their responses from the last failed attempt to reference during the assessment. A two-tailed, independent-samples *t*-test indicated that there was no significant difference in the number of assessment attempts between the Basic ($M = 2.00$, $SD = 0.68$) and Advanced ($M = 1.64$, $SD = 1.03$) Mental Model groups, independent-samples $t(53) = 1.52$, $p = .135$, $d = 0.41$.

Pathfinder Nodes

Pathfinder network analysis requires a list of concepts (or keywords) to serve as nodes in the Pathfinder network structure. This study used eleven nodes that were derived from the training materials. The following is the list of nodes from the current study: *battery management*, *aircraft steering*, *collision avoidance*, *flip battery switches*, *move joystick*, *detect other aircraft*, *heuristic solver*, *target battery level*, *switch temperature*, *autopilot*, and *machine learning*. These nodes correspond to purpose-based (battery management, aircraft steering, collision avoidance), performance-based (flip battery switches, move joystick, detect other aircraft), and process-based (heuristic solver, target battery level, switch temperature) information about the autopilot system. The *autopilot* node was included to relate each concept to the overall system. *Machine learning* was not mentioned in the training slides for either group but was included as a distractor.

Scenarios

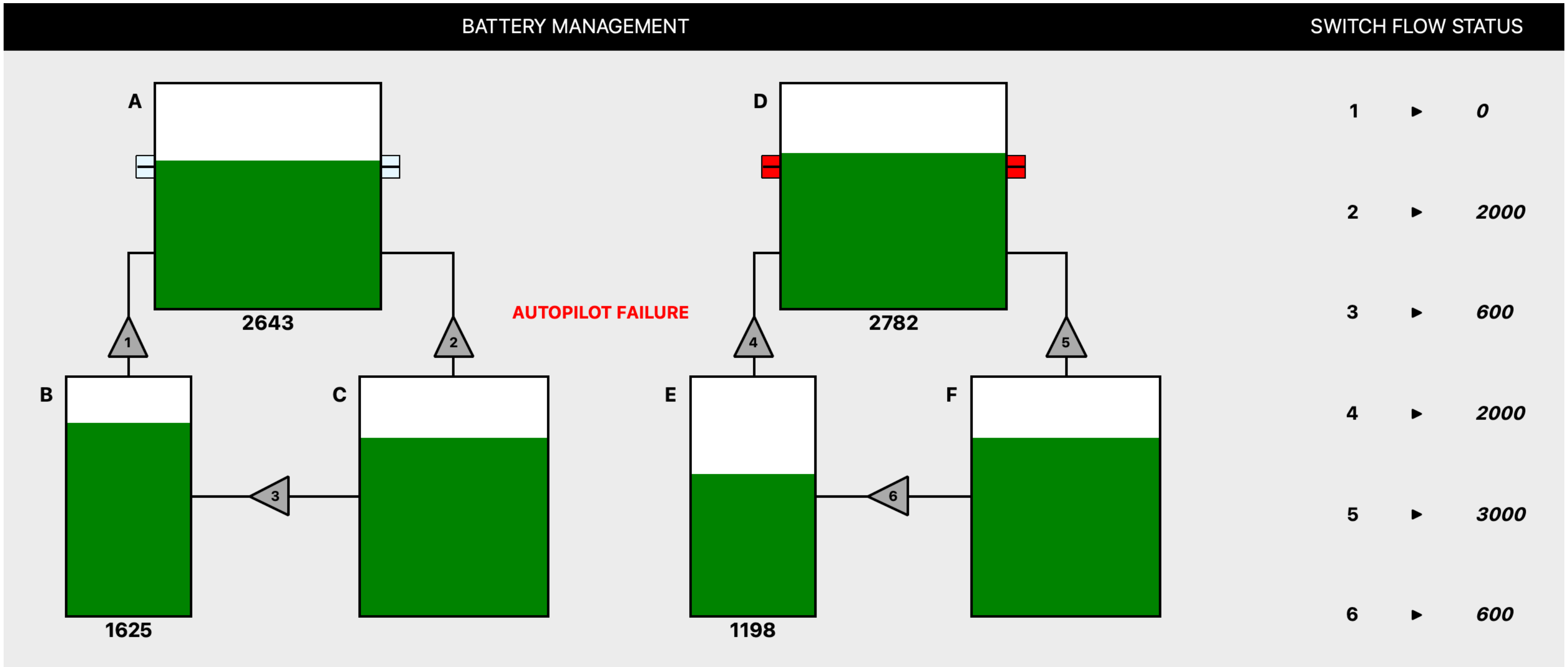
Four similar scenarios were developed with a configurable design that enabled any scenario to be used for either information bandwidth (i.e., within-group) condition. Thus, all participants experienced the same four scenarios, but the scenario order and information bandwidth setting varied by participant. The duration of each scenario was six minutes. The two scenarios for each bandwidth condition (Low, High) were grouped into blocks to allow participants to become calibrated to the bandwidth of the Supplemental Data Display within each block. The order of the blocks was counterbalanced within each group, and the order of the scenarios was randomized for each participant.

Three practice scenarios were created to familiarize participants with the tasks. The first scenario lasted 40 seconds and only included Resource Management Task events. The second scenario also lasted 40 seconds and only included System Monitoring Task events. The final scenario lasted two minutes and included events from both tasks. For all practice scenarios, the Supplemental Data Display bandwidth was set to a frequency (i.e., data refresh rate) that had a data refresh period (600 ms) halfway between the settings for the Low (1000 ms) and High (200 ms) bandwidth conditions. The training scenarios were slightly less difficult than the experimental scenarios based on overall event rate.

For the Resource Management Task (Figure 5), the goal is to maintain appropriate levels of power for the primary battery (i.e., top battery) within each aircraft. Each block represents a battery, and the green color within the block represents the current power level. The two independent groups of batteries in Figure 5 represent two separate aircraft. Automation was always actively controlling the power levels, but participants were instructed to monitor the system for possible failures, which were indicated by an “autopilot failure” alert. If a failure was detected, the participant had to determine which switch (represented by gray triangles in Figure 5) was malfunctioning. A malfunctioning switch prevents the flow of power despite the switch flow status showing a positive flow rate. The participant had to click the malfunctioning switch using the mouse to fix the autopilot failure. If no response was provided by the participant, the autopilot system continued to malfunction until another malfunction occurred within the same aircraft. The error rate for this automation was set to an average of three per minute (i.e., relatively high) to establish an appropriate task difficulty level and to ensure a high level of potential utility for the Supplemental Data Display, which is intended to aid in the resolution of these errors. See training slides in Appendix A for complete task instructions.

Figure 5

Resource Management Task (OpenMATB Software)

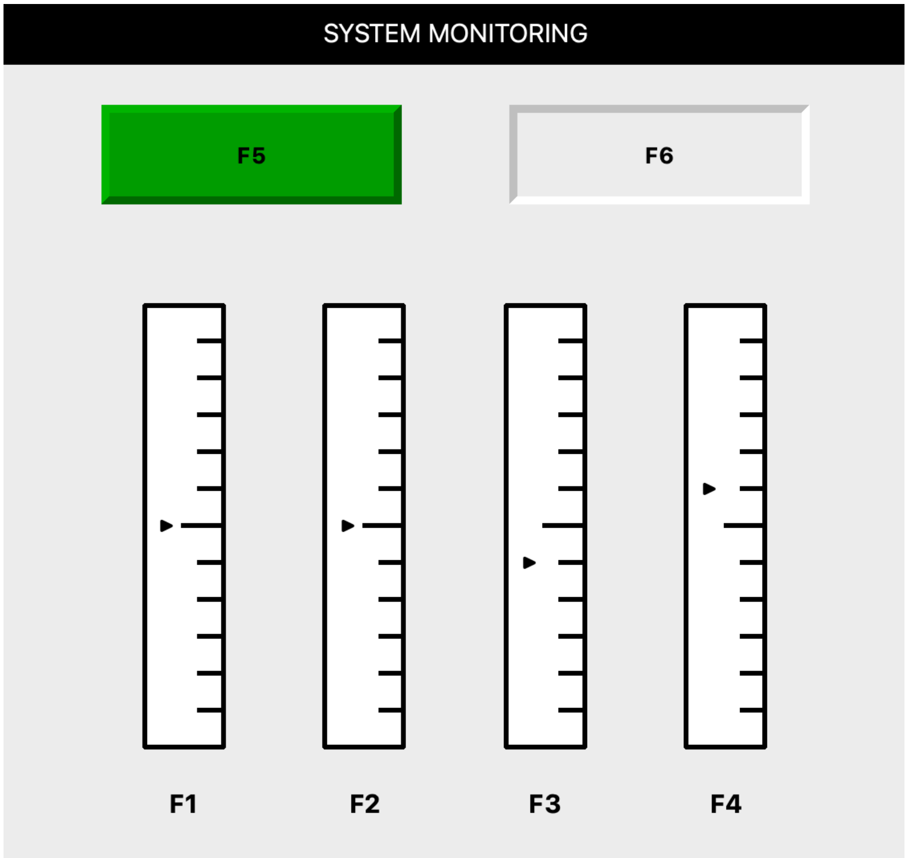


Note. This example shows an autopilot failure in the right aircraft. The behavior of this task was modified from the original version of the Resource Management Task included in the OpenMATB software.

For the System Monitoring Task (Figure 6), the goal is to maintain the normal state of the system, which is completely independent of the Resource Management Task. There are four gauges labeled as F1, F2, F3, and F4 representing the system state. The small black arrow within each gauge fluctuates around the center of the gauge when the system is in a normal state. When a system malfunction occurs, the arrow moves to the top or bottom edge of the gauge. In this case, the participant used the mouse to click the corresponding button for the gauge (F1-F4) to correct the malfunction. Above the gauges are two indicator lights (labeled as F5 and F6) that indicate additional system malfunctions. The green light (F5) ON suggests normal state, and in the event of a specific system malfunction, the green light (F5) turns OFF. The red light (F6) OFF suggests normal state, and in the event of a specific system malfunction, the red light (F6) turns ON. If either light indicated a malfunction, the participant used the mouse to click the corresponding button for the indicator light to correct the system malfunction, which returned the light to normal state.

Figure 6

System Monitoring Task (OpenMATB Software)



Note. All gauges and indicators in this example are shown in a normal state.

The primary task for all scenarios was the Resource Management Task, and the secondary task was the System Monitoring Task. However, the System Monitoring Task had a high rate of events, thus requiring significant attention from the participant. This high demand from the secondary task was intended to encourage the Advanced Mental Model group to utilize the information available in the Supplemental Data Display to reduce the attentional resources required to manage the primary task.

PROCEDURE

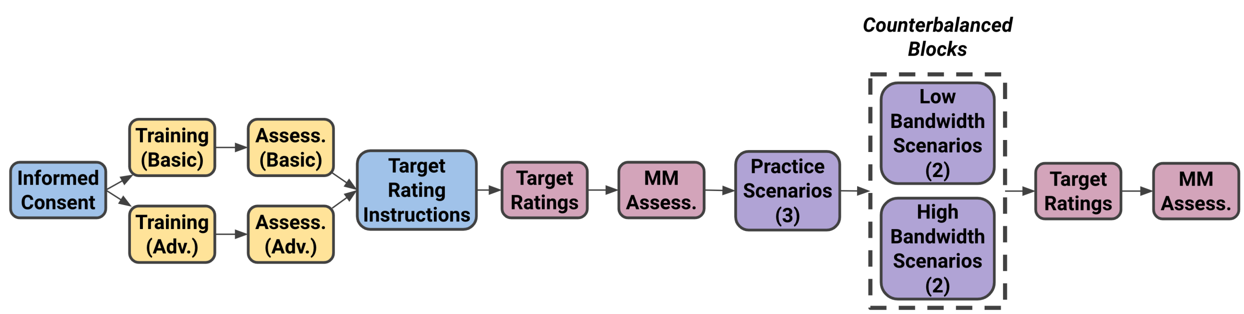
Participants completed the study in a quiet room. Participants first read an informed consent document and indicated whether they agreed to participate in the study. Upon their consent, they completed a demographics questionnaire. Participants then were randomly assigned to either the Basic Mental Model or Advanced Mental Model training group and received their respective training module (see Appendix A). Participants read a training presentation that introduced the concept of Advanced Air Mobility (AAM), explained the experimental task, and described the functionality of the autopilot system present on the vehicles that would be managed in the task. Participants then completed a 12-question training assessment to confirm the training manipulation (see Appendix B). If any questions were answered incorrectly, the participant was asked to review the training materials and retake the assessment. This process was repeated until all questions were answered correctly. The training materials were not available to the participants during the first two attempts at the assessment, but after the second failed attempt, participants were given access to the training slides as well as their responses from the last failed attempt to reference during the assessment. Training (without repeated assessments) took approximately 15 minutes to complete.

Following successful completion of the training assessment, participants completed three practice scenarios, which were simplified versions of the experimental scenarios, to become familiarized with the tasks. After the practice scenarios, participants read instructions on using the target rating tool and then used the tool to provide pairwise relatedness ratings. Participants then completed a 9-question mental model assessment (see Appendix C). Following the assessment, participants ran the four experimental scenarios (blocks of two grouped by information bandwidth condition). The order of the bandwidth blocks was counterbalanced across participants, and the order of the scenarios within each block was randomized.

After completing all of the scenarios, participants again provided pairwise relatedness ratings using the target rating tool and completed the mental model assessment. Participants were then debriefed, thanked for their participation, and dismissed. The study took approximately one and a half hours to complete. The full procedure timeline is presented in Figure 7.

Figure 7

Experiment Procedure Timeline



DEPENDENT VARIABLES

Percentage Dwell Time (PDT)

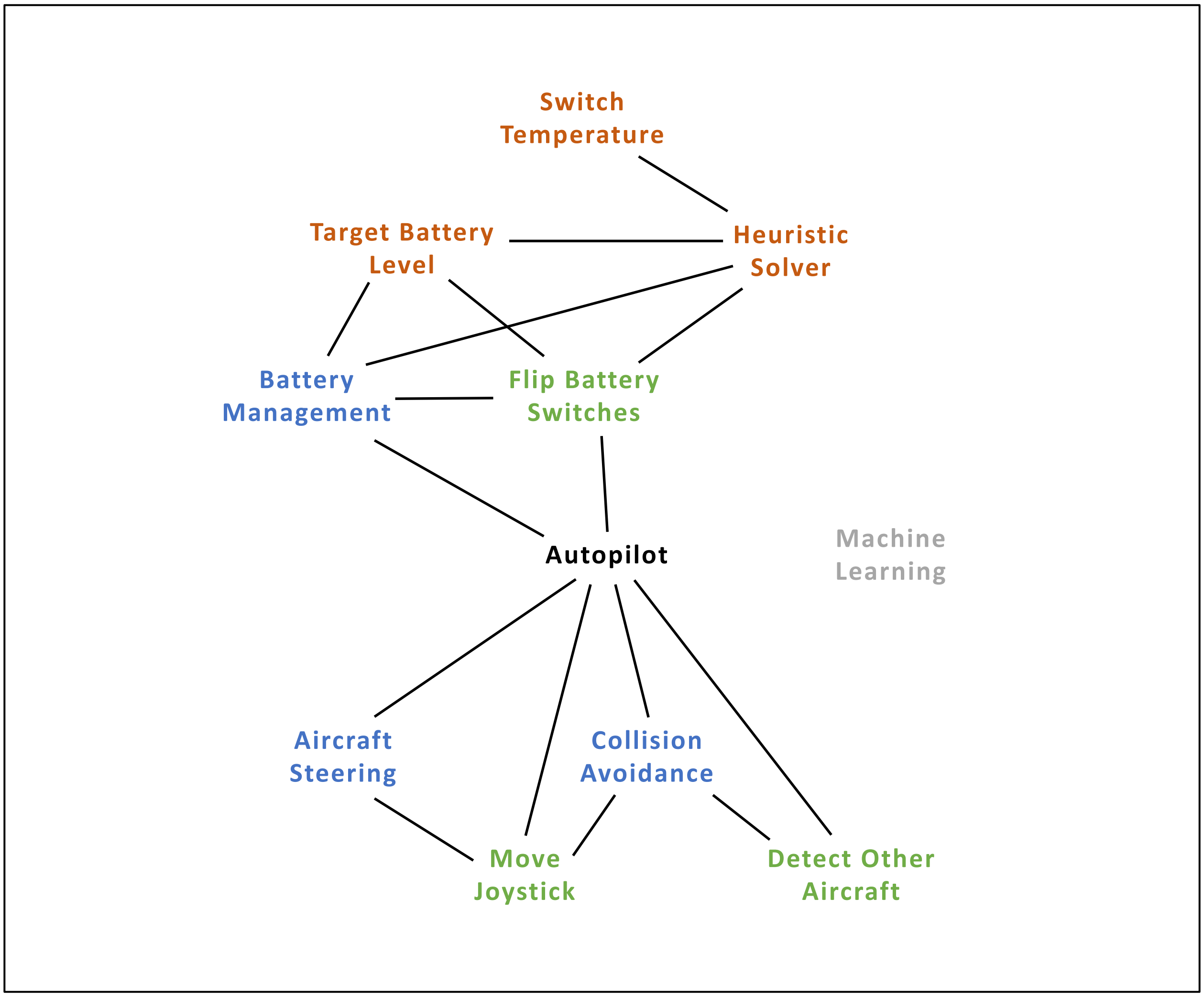
PDT is derived from the eye movement data and refers to the proportion of time that the participant fixated within a specific area of interest (AOI). The AOIs for this experiment are defined as the Resource Management Task window, the System Monitoring Task window, and the Supplemental Data Display window.

Similarity to Design Model (Pathfinder Analysis)

Similarity to the design model value was calculated for each participant, which is the similarity score (i.e., *C* statistic) of the individual network and the design model network. The design model, which is the conceptual model held by the designer (Norman, 1986), was derived from the training materials (Figure 8). This metric is an evaluation of shared links between two networks. The similarity score of two individual networks is calculated using the formula, $X/[T - X]$, where X is the number of common links in the two networks and T is the total number of links across both networks (Goldsmith & Davenport, 1990; Lim & Klein, 2006). This metric ranges on a scale of 0 (completely unrelated) to 1 (identical), so a score of 0.3 implies that 30% of the mental model structure is shared.

Figure 8

Design Model Pathfinder Network



Note. Blue text signifies purpose-based information nodes. Green text signifies performance-based information nodes. Orange text signifies process-based information nodes. Black text signifies the node representing the high-level system. Gray text signifies a distractor node.

Mental Model Assessment Score

A nine-question, multiple-choice mental model assessment (Appendix C) was developed based on the autopilot design model to measure the effect of the mental model manipulation. The design model is the conceptual model held by the designer (Norman, 1986), which is directly depicted in the training materials. The questions were developed based on the training materials to assess purpose-based, process-based, and performance-based information about the autopilot (three questions for each type of information).

Coherence and Internal Consistency (Pathfinder Analysis)

Coherence and internal consistency were calculated for each individual network. Coherence is a measure of the reliability of an individual's relatedness ratings, as determined by the transitive consistency of those ratings (Interlink, 2017a). For example, if a participant's ratings indicate that Concept A is similar to Concept B, and Concept B is similar to Concept C, then Concept A should also be similar to Concept C. A higher coherence value for a participant's set of ratings indicates that these transitive relationships are more reliable; thus, the complete set of ratings is more coherent. The range of this metric is -1 (inverse coherence) to 1 (perfect coherence).

Internal consistency is a binary value derived from the coherence value to determine the overall reliability of an individual Pathfinder network. Coherence greater than or equal to 0.15 is considered adequate internal consistency, whereas coherence less than 0.15 may indicate a poor understanding of the concepts being rated (Interlink, 2017a). If coherence values are on average lower than 0.15, the Pathfinder data should not be analyzed.

Resource Management Task Performance

Performance in the Resource Management Task was measured by the hit rate, the number of false alarms, the response time (RT) for hit events, and the cumulative amount of time that the power level was outside of the acceptable range. The RT for a hit event is defined as the time elapsed from malfunction onset to participant response. A false alarm event occurs if the participant responds to a unit that is not malfunctioning. Because a false alarm event implies no malfunction onset, RT cannot be calculated for a false alarm event.

System Monitoring Task Performance

Performance in the System Monitoring Task was measured by the hit rate, the number of false alarms, and the RT for hit events. The RT for a hit event is defined as the time elapsed from signal onset to participant response. A false alarm event occurs if the participant responds when a signal is not present. Because a false alarm event implies no signal onset, RT cannot be calculated for a false alarm event.

CHAPTER III

RESULTS

The following analyses were conducted using 2 × 2 split-plot factorial analyses of variance (ANOVAs). For each variable, the assumption of normality was checked by observation of histograms of values for each condition. Checks indicated that data were generally unimodal and normally distributed for all variables included in this analysis except for several that were skewed due to floor or ceiling effects. However, ANOVA is generally robust to violations of the normality assumption, so the analysis is still appropriate for the present data set (Maxwell & Delaney, 2004). Levene's test was conducted for each between-subjects measure to test the assumption of homogeneity of variance. Some measures included in the following analyses violated this assumption, and results of the Levene's test are reported for those violations. ANOVA is generally robust to moderate violations of the assumption of homogeneity of variance when the sample sizes of groups are approximately equal and group sample sizes are not unreasonably small (e.g., less than five per group; Maxwell & Delaney, 2004). Thus, ANOVA is still an appropriate analysis to conduct on the present data despite violations of this assumption. The assumption of sphericity is not applicable for this analysis because the within-subjects factors only include two conditions. An alpha level of $p < .05$ was established to indicate statistical significance. Bonferroni corrections (i.e., dividing the alpha by the number of comparisons) were used for all post-hoc analyses.

MENTAL MODEL ASSESSMENTS

The *Similarity to Design Model (Pathfinder Analysis)* and *Mental Model Assessment* scores were used to measure the effect of the mental model manipulation in the current study. For all Pathfinder network analyses, Pathfinder (Interlink, 2021) software was used with parameter settings $r = \infty$ and $q = 10$. Furthermore, Pathfinder network analysis requires only a single rating score for each pair of nodes, but the Target Rating method collects two ratings per node pair. For the current study, the duplicate ratings (within each single set of target ratings) were averaged to provide a single rating score for each pair (see Politowicz et al., 2022).

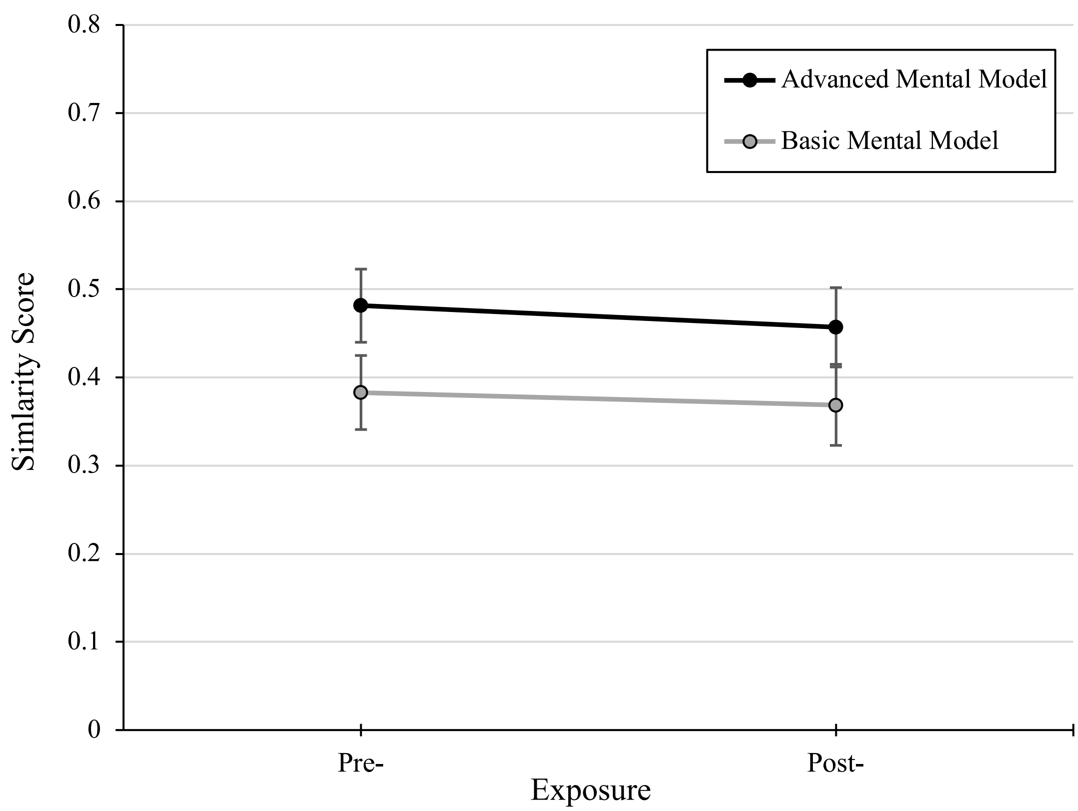
Prior to conducting the Pathfinder network similarity analyses, the mean *Coherence* value for each mental model group was calculated to determine if the Pathfinder network data met the threshold for internal consistency (i.e., mean *Coherence* greater than or equal to 0.15). The mean coherence values exceeded the threshold for internal consistency for both the Basic Mental Model group (Pre-exposure: $M = 0.75$, $SD = 0.10$; Post-exposure: $M = 0.70$, $SD = 0.16$) and Advanced Mental Model group (Pre-exposure: $M = 0.74$, $SD = 0.13$; Post-exposure: $M = 0.72$, $SD = 0.15$). Therefore, the Pathfinder data were found to be suitable for further analysis.

To assess if participants' mental model similarity to design model scores (Pathfinder analysis) were affected by mental model condition (Basic vs. Advanced) and exposure to the system (Pre-exposure vs. Post-exposure), a 2 × 2 split-plot factorial ANOVA was conducted. Levene's test indicated a violation of the assumption of homogeneity of variance for the Post-exposure condition, $F(1, 53) = 4.96$, $p = .030$, but ANOVA is still considered an appropriate analysis for the data set despite this violation (Maxwell & Delaney, 2004).

The ANOVA revealed that there was a significant main effect of mental model condition, such that the Advanced Mental Model group networks ($M = 0.47$) were significantly more similar to the design model than were the Basic Mental Model group networks ($M = 0.38$), $F(1, 53) = 11.35$, $p = .001$, $\eta_p^2 = .176$. The main effect of exposure was not significant, $F(1, 53) = 2.18$, $p = .146$, $\eta_p^2 = .040$, and there was no significant interaction between mental model condition and exposure, $F(1, 53) = 0.16$, $p = .690$, $\eta_p^2 = .003$. Figure 9 shows the mean Pathfinder analysis similarity score for all conditions. Descriptive statistics for all conditions are included in Table 1. Mental model aggregate networks for both groups and both exposure conditions are shown in Figure 10.

Figure 9

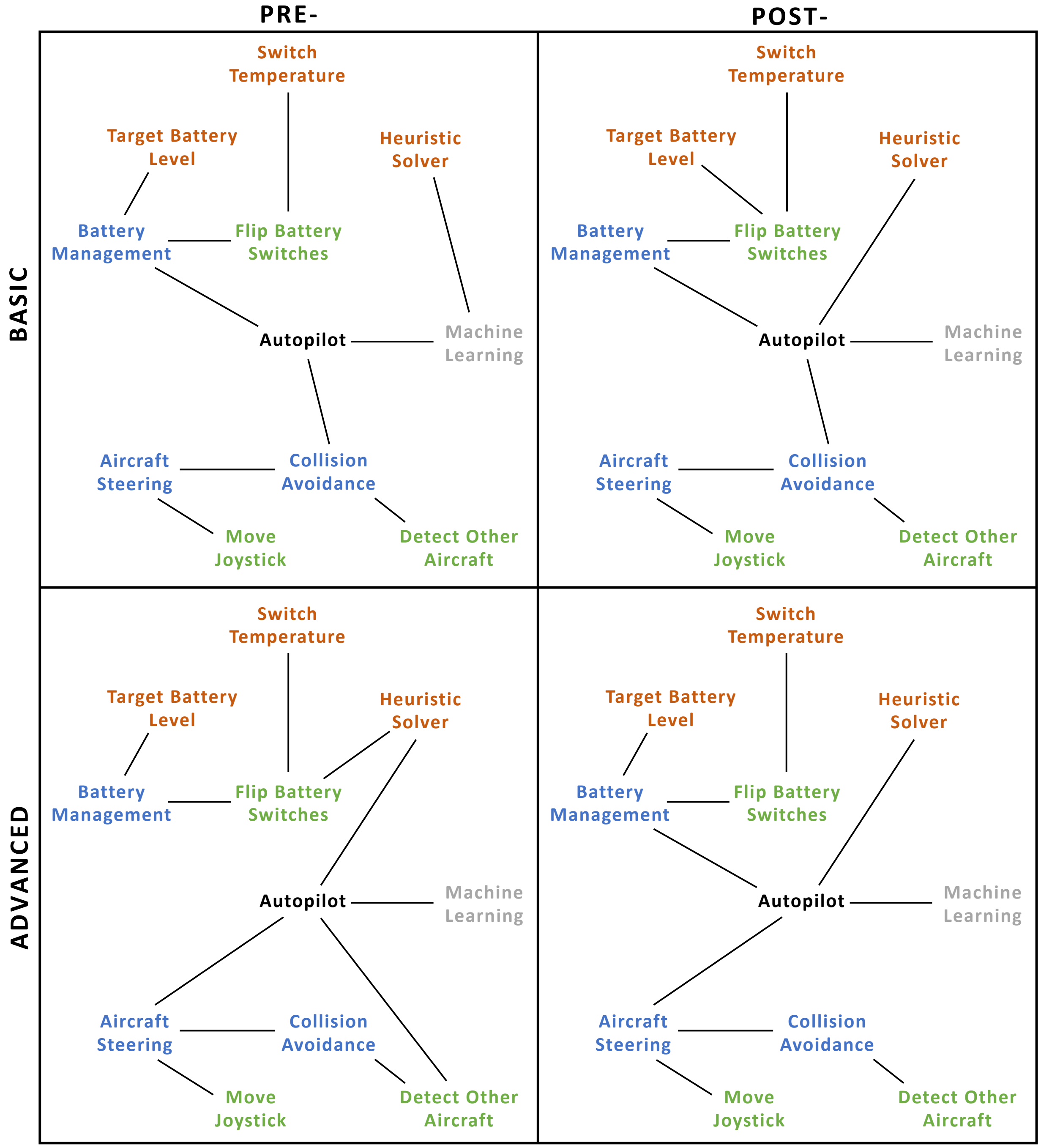
Pathfinder Similarity Scores by Mental Model Group and Exposure



Note. Error bars represent between-subjects 95% confidence intervals.

Figure 10

Pathfinder Mean Aggregate Networks by Mental Model Group and Exposure

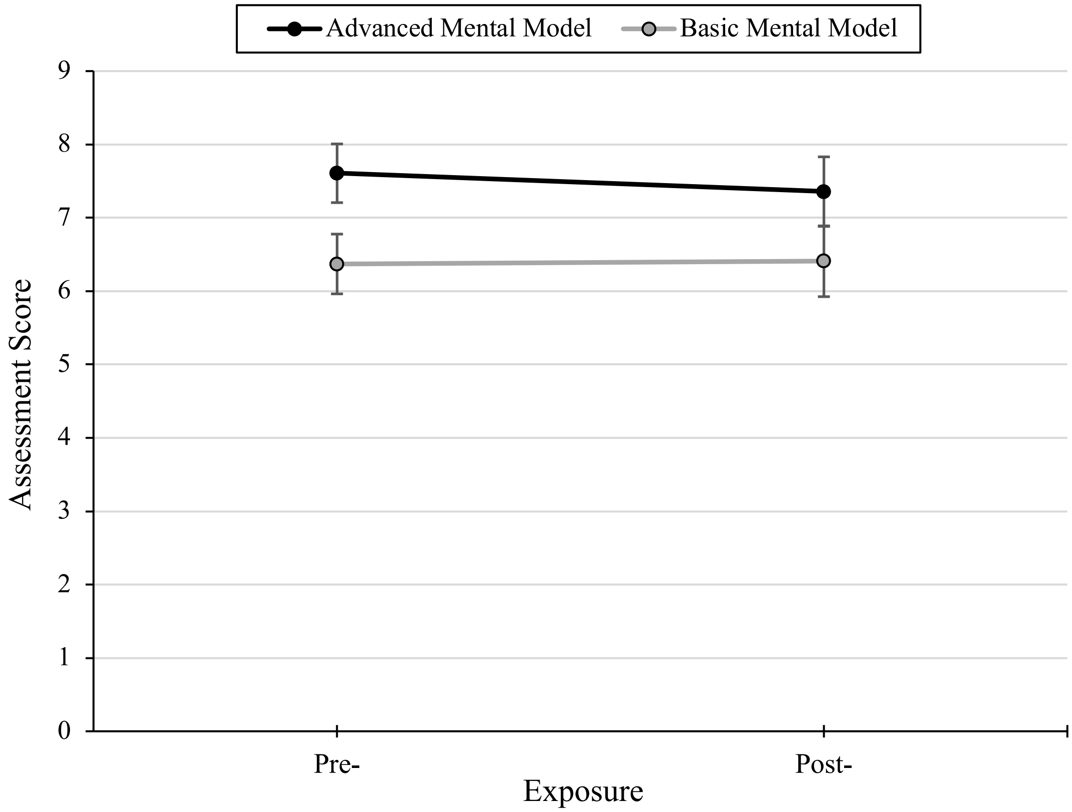


Note. Blue text signifies purpose-based information nodes. Green text signifies performance-based information nodes. Orange text signifies process-based information nodes. Black text signifies the node representing the high-level system. Gray text signifies a distractor node.

To assess if participants' mental model assessment scores were affected by mental model condition (Basic vs. Advanced) and exposure to the system (Pre-exposure vs. Post-exposure), a 2 × 2 split-plot factorial ANOVA was conducted. The ANOVA revealed that there was a significant main effect of mental model condition, such that the Advanced Mental Model group (M = 7.48) scored significantly higher than the Basic Mental Model group (M = 6.39), $F(1, 53) = 14.18$, $p < .001$, $\eta_p^2 = .211$. The main effect of exposure was not significant, $F(1, 53) = 0.87$, $p = .357$, $\eta_p^2 = .016$, and there was no significant interaction between mental model condition and exposure, $F(1, 53) = 1.57$, $p = .216$, $\eta_p^2 = .029$. Figure 11 shows the mean mental model assessment score for all conditions. Descriptive statistics for all conditions are included in Table 1.

Figure 11

Mental Model Assessment Scores by Mental Model Group and Exposure



Note. Error bars represent between-subjects 95% confidence intervals.

Table 1

Descriptive Statistics for Mental Model Measures

Variable	*n*	*Similarity (Pathfinder)*	*Mental Model Assessment*
Basic Mental Model			
Pre-exposure	27	0.38 (0.09)	6.37 (0.97)
Post-exposure	27	0.37 (0.10)	6.41 (1.34)
Advanced Mental Model			
Pre-exposure	28	0.48 (0.12)	7.61 (1.13)
Post-exposure	28	0.46 (0.13)	7.36 (1.16)

Note. Data presented in the table are mean values with standard deviations in parentheses.

PRIMARY TASK PERFORMANCE

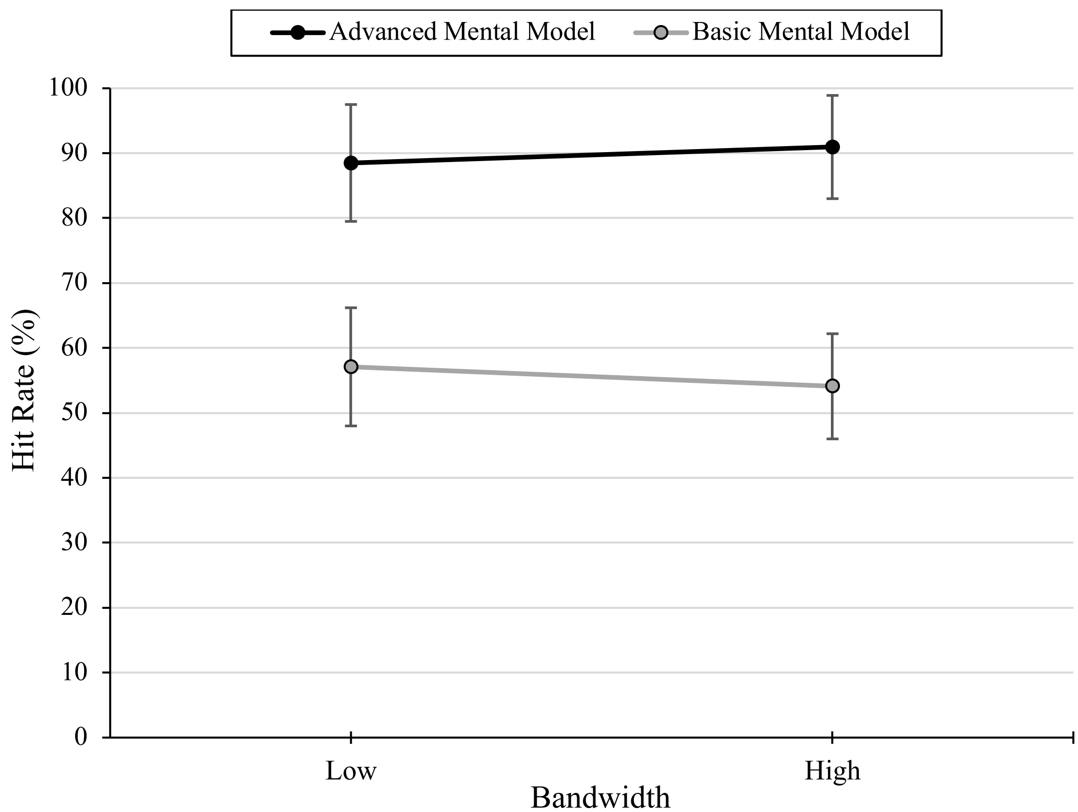
To assess if participants' hit rate, hit RT, number of false alarms, and cumulative time outside of acceptable battery range for the Resource Management Task (primary task) were affected by mental model condition (Basic vs. Advanced) and information bandwidth (Low vs. High), 2 × 2 split-plot ANOVAs were conducted. Descriptive statistics for all performance measures and conditions are shown in Table 2.

For hit rate in the Resource Management Task, the ANOVA revealed that there was a significant main effect of mental model condition, such that the Advanced Mental Model group (M = 89.7%) had significantly higher hit rates than the Basic Mental Model group (M = 55.6%),

$F(1, 53) = 36.89$, $p < .001$, $\eta_p^2 = .410$. The main effect of bandwidth was not significant, $F(1, 53) = 0.01$, $p = .909$, $\eta_p^2 = .000$, and there was no significant interaction between mental model condition and bandwidth, $F(1, 53) = 1.55$, $p = .219$, $\eta_p^2 = .028$. Figure 12 shows the mean hit rates for all conditions.

Figure 12

Primary Task Hit Rate by Mental Model Group and Bandwidth



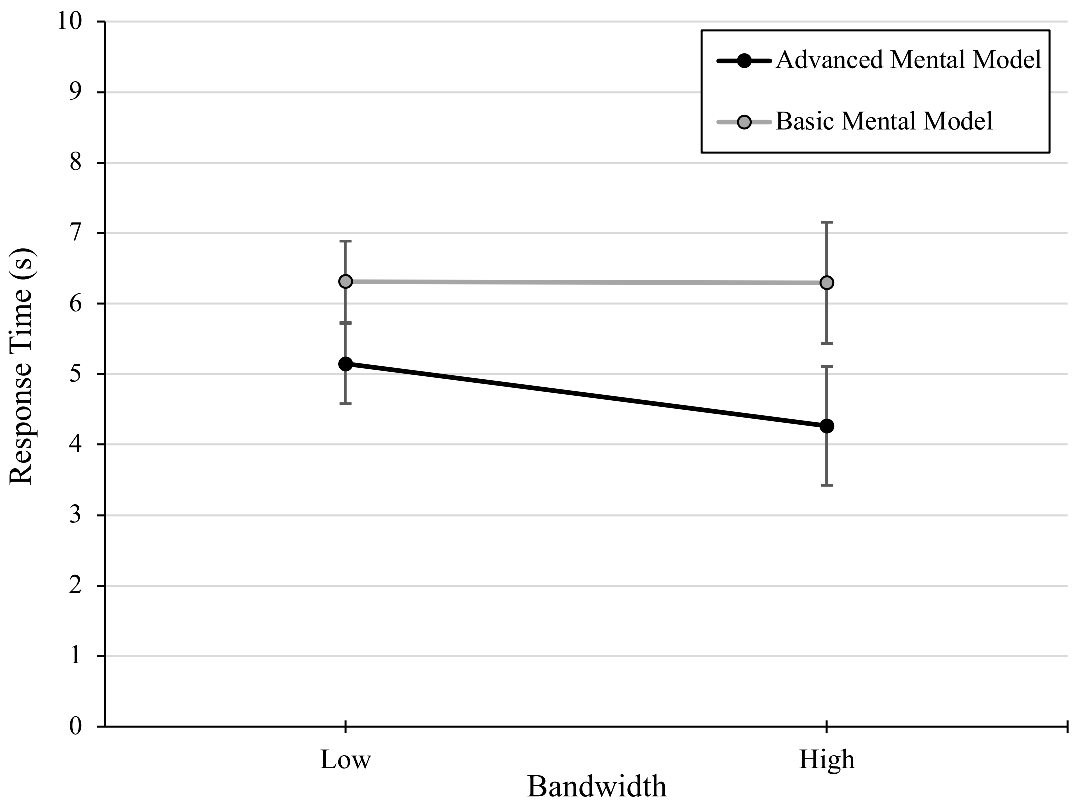
Note. Error bars represent between-subjects 95% confidence intervals.

For hit RT in the Resource Management Task, the ANOVA revealed that there was a significant main effect of mental model condition, such that the Advanced Mental Model group ($M = 4.71$ seconds) had significantly faster RTs than the Basic Mental Model group ($M = 6.30$

seconds), $F(1, 53) = 12.49$, $p < .001$, $\eta_p^2 = .191$. Therefore, there is no evidence for speed-accuracy tradeoffs. The main effect of bandwidth was not significant, $F(1, 53) = 3.50$, $p = .067$, $\eta_p^2 = .062$, and there was no significant interaction between mental model condition and bandwidth, $F(1, 53) = 3.25$, $p = .077$, $\eta_p^2 = .058$. Figure 13 shows the mean hit RTs for all conditions.

Figure 13

Primary Task Hit Response Time (RT) by Mental Model Group and Bandwidth



Note. Error bars represent between-subjects 95% confidence intervals.

For number of false alarms in the Resource Management Task, the ANOVA revealed that there was no significant main effect of mental model condition, $F(1, 53) = 2.68$, $p = .107$, $\eta_p^2 =$

.048, and no significant main effect of bandwidth, $F(1, 53) = 0.15$, $p = 0.702$, $\eta_p^2 = .003$. There was also no significant interaction between mental model condition and bandwidth, $F(1, 53) = 0.48$, $p = .494$, $\eta_p^2 = .009$.

For cumulative time outside of the acceptable battery range in the Resource Management Task, the ANOVA revealed that there was a significant main effect of mental model condition, such that the Advanced Mental Model group ($M = 130.96$ seconds) spent significantly less time outside of the acceptable range than the Basic Mental Model group ($M = 232.78$ seconds), $F(1, 53) = 14.90$, $p < .001$, $\eta_p^2 = .219$. There was also a significant main effect of bandwidth, such that the High Bandwidth condition ($M = 167.85$ seconds) spent significantly less time outside of the acceptable range than the Low Bandwidth condition ($M = 195.88$ seconds), $F(1, 53) = 4.93$, $p = .031$, $\eta_p^2 = .085$. There was no significant interaction between mental model condition and bandwidth, $F(1, 53) = 0.86$, $p = .357$, $\eta_p^2 = .016$.

Table 2

Descriptive Statistics for Primary Task Performance Measures

Variable	*n*	*Hit Rate (%)*	*Hit RT (s)*	*Number of False Alarms*	*Time Out of Range (s)*
Basic Mental Model					
Low Bandwidth	27	57.10 (23.12)	6.31 (1.48)	8.59 (11.11)	240.93 (112.02)
High Bandwidth	27	54.12 (20.45)	6.29 (1.72)	7.96 (10.57)	224.62 (93.68)
Advanced Mental Model					
Low Bandwidth	28	88.49 (24.20)	5.15 (1.51)	3.66 (8.84)	150.83 (119.75)
High Bandwidth	28	90.97 (21.47)	4.27 (2.62)	3.84 (11.24)	111.08 (106.18)

Note. Data presented in the table are mean values with standard deviations in parentheses.

SECONDARY TASK PERFORMANCE

To assess if participants' hit rate, hit RT, and number of false alarms for the System Monitoring Task (secondary task) were affected by mental model condition (Basic vs. Advanced) and information bandwidth (Low vs. High), 2 × 2 split-plot ANOVAs were conducted. The ANOVAs revealed that there were no significant main effects or interactions for any of the secondary task performance measures, indicating there are no observable changes in performance tradeoffs between the primary Resource Management Task and the secondary System Monitoring Task. Descriptive statistics for all performance measures and conditions are shown in Table 3.

For hit rate in the System Monitoring Task, the ANOVA revealed that there was not a significant main effect of mental model condition, $F(1, 53) = 0.25$, $p = .617$, $\eta_p^2 = .005$, or bandwidth, $F(1, 53) = 0.01$, $p = .906$, $\eta_p^2 = .000$. There was also no significant interaction between mental model condition and bandwidth, $F(1, 53) = 0.03$, $p = .858$, $\eta_p^2 = .001$.

For hit RT in the System Monitoring Task, the ANOVA revealed that there was not a significant main effect of mental model condition, $F(1, 53) = 0.02$, $p = .900$, $\eta_p^2 = .000$, or bandwidth, $F(1, 53) = 2.33$, $p = .133$, $\eta_p^2 = .042$. There was also no significant interaction between mental model condition and bandwidth, $F(1, 53) = 0.90$, $p = .347$, $\eta_p^2 = .017$.

For number of false alarms in the System Monitoring Task, the ANOVA revealed that there was not a significant main effect of mental model condition, $F(1, 53) = 0.08$, $p = .777$, $\eta_p^2 = .002$, or bandwidth, $F(1, 53) = 0.03$, $p = .856$, $\eta_p^2 = .001$. There was also no significant interaction between mental model condition and bandwidth, $F(1, 53) = 0.41$, $p = .524$, $\eta_p^2 = .008$.

Table 3

Descriptive Statistics for Secondary Task Performance Measures

Variable	*n*	*Hit Rate (%)*	*Hit RT (s)*	*Number of False Alarms*
Basic Mental Model				
Low Bandwidth	27	84.57 (23.69)	2.33 (0.56)	2.06 (3.13)
High Bandwidth	27	84.65 (27.03)	2.29 (0.43)	1.70 (3.77)
Advanced Mental Model				
Low Bandwidth	28	87.80 (18.84)	2.37 (0.72)	2.07 (4.27)
High Bandwidth	28	87.43 (20.25)	2.20 (0.70)	2.27 (4.94)

Note. Data presented in the table are mean values with standard deviations in parentheses.

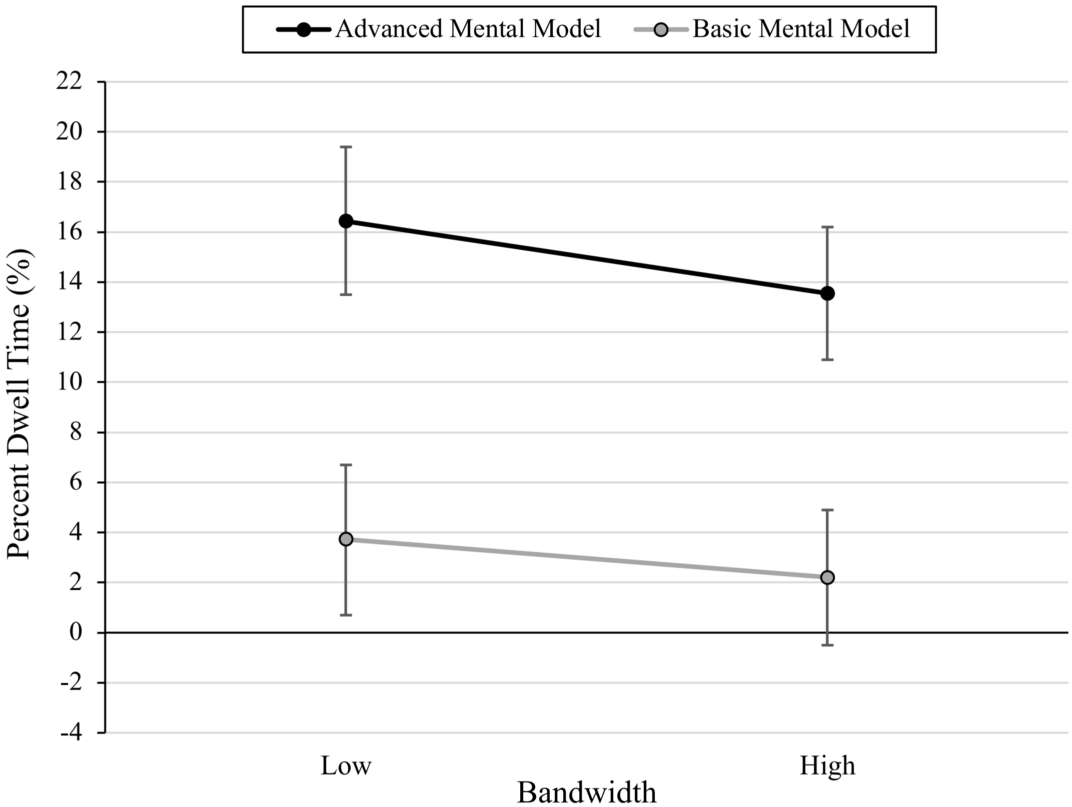
PDT ON SUPPLEMENTAL DATA DISPLAY AOI

To assess if participants' PDT on the Supplemental Data Display AOI was affected by mental model condition (Basic vs. Advanced) and information bandwidth (Low vs. High), a 2 × 2 split-plot factorial ANOVA was conducted. Levene's test indicated violations of the assumption of homogeneity of variance for the Low Bandwidth condition, $F(1, 53) = 5.30$, $p = .025$, and the High Bandwidth condition, $F(1, 53) = 9.36$, $p = .003$. However, ANOVA is still considered an appropriate analysis for the data set despite this violation (Maxwell & Delaney, 2004).

The ANOVA revealed that there was a significant main effect of mental model condition, such that the Advanced Mental Model group (M = 15.0%) had significantly higher PDTs than the Basic Mental Model group (M = 3.0%), $F(1, 53) = 43.48$, $p < .001$, $\eta_p^2 = .451$. There was also a significant main effect of bandwidth, such that the Low Bandwidth condition (M = 10.1%) had significantly higher PDTs than the High Bandwidth condition (M = 7.9%), $F(1, 53) = 7.79$, $p = .007$, $\eta_p^2 = .128$. Unexpectedly, however, there was no significant interaction between mental model condition and bandwidth, $F(1, 53) = 0.76$, $p = .389$, $\eta_p^2 = .014$. Figure 14 shows the mean PDT on the Supplemental Data Display AOI for all conditions. Descriptive statistics for all conditions are included in Table 4.

Figure 14

PDT on Supplemental Data Display AOI by Mental Model Group and Bandwidth



Note. Error bars represent between-subjects 95% confidence intervals.

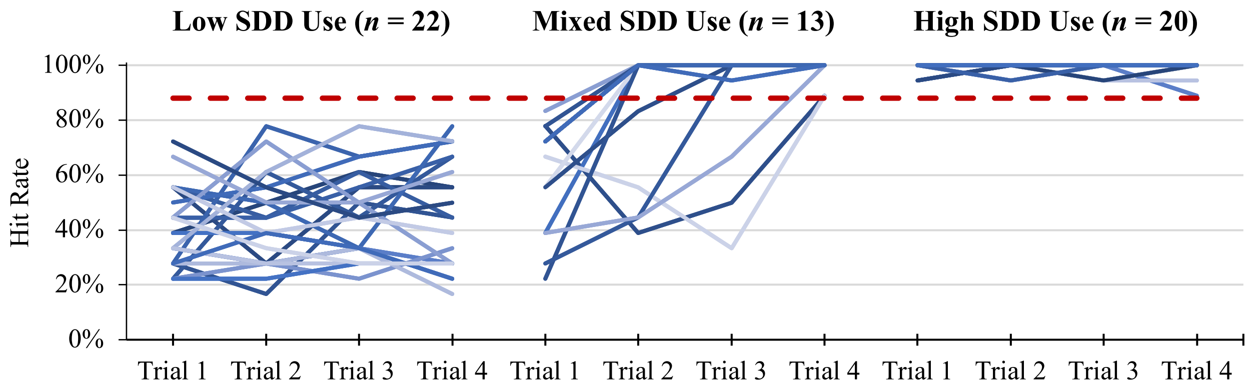
EVALUATION OF TASK VALUE GROUPS

To evaluate the relationship more rigorously between mental model group and information bandwidth, the PDT data was further analyzed with more stringent exclusion criteria within each mental model group based on performance data. The functional behavior of the Resource Management Task enables the inference of Supplemental Data Display (SDD) information usage from user performance data, specifically the hit rate. When a user does not utilize the SDD information, the user is often required to guess which switch has failed because in many instances, there is no useful information available to assist with the decision. Thus, a participant's maximum hit rate for the task is essentially limited without the SDD information. Through pilot testing and analysis of the present data set, an 88.89% hit rate (2 misses out of 18 failures) was found to be the appropriate threshold for determining if a user has utilized the SDD information. Though it is statistically possible on a given trial for a participant who is not using the SDD information to meet or exceed this threshold, it is highly unlikely that this would occur across all four trials. Thus, the criteria for determining low (or no) SDD use is when participant scores are lower than 88.89% hit rate across all four trials (Figure 15). Similarly, although it is possible for a participant using the SDD information to have a hit rate below the threshold, this was found to be rare because use of the SDD information leads to perfect or near-perfect performance with little effort. Thus, the criteria for determining high SDD use is when participant scores are equal to or higher than 88.89% hit rate across all four trials (Figure 15).

In the present study, a subset of participants scored below the threshold in some trials and above in other trials for various reasons, such as discovering the relevance of the SDD information during a trial. Therefore, any participant who did not score fully above or fully below the 88.89% hit rate threshold across all four trials is categorized as having mixed SDD use (Figure 15). Figure 15 shows the categorization of participants across both mental model groups, whereas Figure 16 shows the categorization of participants within each mental model group.

Figure 15

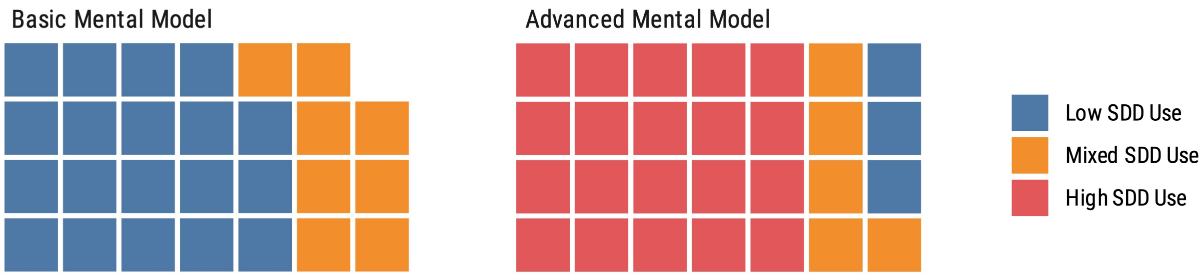
Performance Categories Based on Primary Task Hit Rate Across Trials



Note. Each line represents an individual participant. The figure shows the categorization of all participants from both mental model groups into three performance-based groups. The *y*-axis on the chart depicts the Resource Management Task (primary task) hit rate. Trials are ordered in the order of appearance. SDD = Supplemental Data Display.

Figure 16

Performance Category Distribution of Participants by Mental Model Group



Note. Each square represents an individual participant. The Low Task Value group is comprised of the 19 participants in the Basic Mental Model group who are categorized as low SDD use (blue squares). The High Task Value group is comprised of the 20 participants in the Advanced Mental Model group who are categorized as high SDD use (red squares). SDD = Supplemental Data Display.

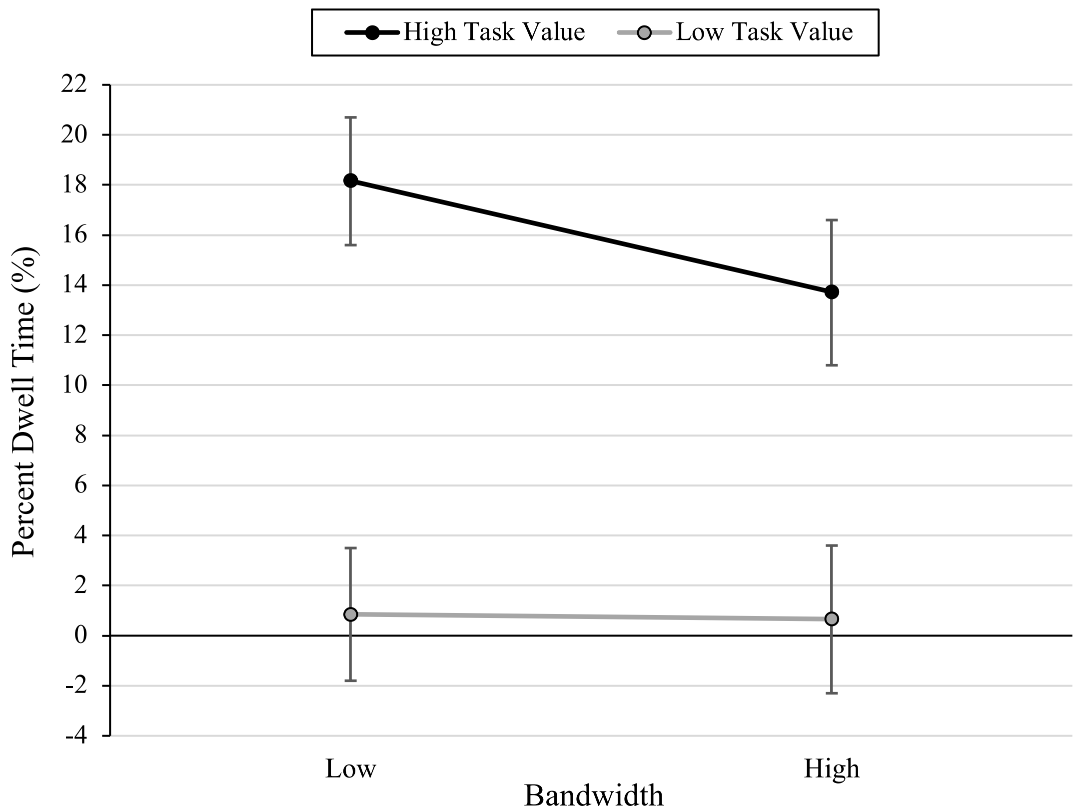
Following the additional criteria, two new groups were established, which are each a subset of one of the mental model groups. The High Task Value group (n = 20) includes participants from the Advanced Mental Model group whose performance data indicates high SDD use (i.e., greater than or equal to 88.89% hit rate across all trials) and therefore high task relevance of the SDD AOI. The Low Task Value group (n = 19) includes participants from the Basic Mental Model group whose performance indicates low SDD use (i.e., less than 88.89% hit rate across all trials) and therefore low task relevance of the SDD AOI. Figure 16 shows the overall categorization of participants based on performance within each mental model group.

To assess if participants' PDT on the Supplemental Data Display AOI was affected by task value group (Low vs. High) and information bandwidth (Low vs. High), a 2 × 2 split-plot factorial ANOVA was conducted. Levene's test indicated violations of the assumption of homogeneity of variance for the Low Bandwidth condition, $F(1, 37) = 21.45$, $p < .001$, and the High Bandwidth condition, $F(1, 37) = 9.93$, $p = .003$. However, ANOVA is still considered an appropriate analysis for the data set despite this violation (Maxwell & Delaney, 2004).

The ANOVA revealed that there was a significant interaction between task value group and bandwidth, $F(1, 37) = 14.37$, $p < .001$, $\eta_p^2 = .280$. Post-hoc *t*-tests (using a Bonferroni-corrected alpha level of $p < .025$) indicated that the high-bandwidth displays had lower PDTs than the low-bandwidth displays for the High Task Value group, indicating that the displays with higher bandwidths attracted less attention than those with lower bandwidths, $M = 13.73\%$ vs. 18.17%, paired-samples $t(19) = 4.10$, $p < .001$, $d = 0.92$. On the other hand, for the Low Task Value group, observed PDT values were statistically indistinguishable, $M = 0.66\%$ vs. 0.85%, paired-samples $t(18) = 1.30$, $p = .210$, $d = 0.30$. Additionally, there was a significant main effect of task value condition, such that the High Task Value group ($M = 15.9\%$) had significantly higher PDTs than the Low Task Value group ($M = 0.8\%$), $F(1, 37) = 67.79$, $p < .001$, $\eta_p^2 = .647$. There was also a significant main effect of bandwidth, such that the Low Bandwidth condition ($M = 9.5\%$) had significantly higher PDTs than the High Bandwidth condition ($M = 7.2\%$), $F(1, 37) = 17.09$, $p < .001$, $\eta_p^2 = .316$. Figure 17 shows the mean PDT on the Supplemental Data Display AOI for all conditions. Descriptive statistics for all conditions are included in Table 4.

Figure 17

PDT on Supplemental Data Display AOI by Task Value Group and Bandwidth



Note. Error bars represent between-subject 95% confidence intervals.

Table 4

Descriptive Statistics for PDT on Supplemental Data Display AOI

	Full Groups		Task Value Groups	
Variable	*n*	*PDT (%)*	*n*	*PDT (%)*
Basic Mental Model				
Low Bandwidth	27	3.73 (5.58)	19	0.85 (0.74)
High Bandwidth	27	2.21 (2.95)	19	0.66 (0.64)
Advanced Mental Model				
Low Bandwidth	28	16.43 (9.38)	20	18.17 (7.79)
High Bandwidth	28	13.55 (9.30)	20	13.73 (8.91)

Note. Data presented in the table are mean values with standard deviations in parentheses.

CHAPTER IV

DISCUSSION

The purpose of this study was to examine whether a manipulation of the mental model of an automated system would influence visual attention allocation in a simulated AAM resource management task with a supplementary data display that updated at different information bandwidths. This study aimed to investigate the relationship between expectancy (i.e., information bandwidth) and value (i.e., mental model) parameters of the SEEV model that assumed either an additive (Wickens, 2015, 2021; Wickens & McCarley, 2008) or interactive (Wickens et al., 2001, 2008, 2022) relationship previously. Participants acted as remote aircraft operators (i.e., pilots) managing multiple vehicles in a low-fidelity flight simulation environment. During a continuous dual-task scenario, participants were responsible for a resource management task (primary task) and a system monitoring task (secondary task). Supplemental autopilot status was presented in a supplemental data display, but only half of the participants were provided with additional process-based information about the autopilot (Advanced Mental Model group) that afforded them the necessary context for the value of the displayed data. The information bandwidth (i.e., display update rate) for the Supplemental Data Display was manipulated between low (1 Hz) and high (5 Hz) settings across trial blocks as a within-subjects manipulation to examine the effect of expectancy.

MENTAL MODELS AND VISUAL ATTENTION

The results of the current study indicate that a direct mental model manipulation led to differences in visual attention allocation as predicted by the SEEV model, which directly supports the first hypothesis. Specifically, the group that was provided with additional process-based information about the autopilot (Advanced Mental Model group) allocated more visual attention towards the Supplemental Data Display than the group that did not receive this information. Importantly, the mental model groups were provided with identical task-based information (e.g., task instructions, task strategy), so the observed differences in visual attention allocation (as measured by PDT on the SDD AOI) resulted from inferences made by participants to establish task-relevant value of the displayed information based on knowledge of the underlying automated system.

The value parameter in the SEEV model is influenced by task relevance and task priority relative to the information provided within an AOI (Wickens et al., 2022). It was hypothesized that mental models of the automated system could influence task relevance (and thus the value parameter). The study employed constant task priority across all conditions to isolate the effect of task relevance, and the training manipulation intentionally avoided explicitly introducing differences in task-based information (e.g., task instructions). Results indicated that there was no significant difference in secondary task performance between mental model groups, which supports the assertion that task priority remained constant across conditions. Thus, the results of the current study suggest that the value parameter of the SEEV model is sensitive to differences in mental models of an automated system used during a task, adding to previous findings that have primarily shown the value parameter is sensitive to differences in expertise (i.e., novice, expert) and differences in mental models of the tasks (i.e., task priorities).

For the Pathfinder mental model assessment, a significant difference between mental model groups was expected, such that the Advanced Mental Model group would have significantly greater similarity to the design model than the Basic Mental Model group for both the pre- and post-exposure assessments. This analysis was intended to confirm the effectiveness of the mental model training manipulation. The results indicated that the training manipulation was in fact effective in altering participants' mental models of the automated system between groups. Furthermore, the Pathfinder networks provide additional insight regarding the differences between mental model groups, reinforcing the effectiveness and utility of Pathfinder for the analysis of mental models of automated systems using a relatively small set of nodes. These results collectively show that the mental model manipulation effectively fine-tuned participants' mental models to assimilate the design model. It is also noteworthy that process-based information from only four training slides substantially influenced the participants' construction of mental models, which caused reliable differences in their allocation of visual attention. These changes in visual scanning patterns presumably influence how operators use and trust the automated system because operators with a more advanced mental model possess a better representation of the function of the system (e.g., Craik, 1943), allowing them to effectively sample information necessary for executing a given task jointly with the automated system (Karpinsky et al., 2018). Practically, automation designers should be mindful that simple instruction, as few as four training slides, can lead to substantial differences in operators' scanning patterns and potentially their use, or misuse, of the automation (Parasuraman & Riley, 1997).

EXPECTANCY AND VISUAL ATTENTION

The information bandwidth manipulation also caused changes in visual attention allocation, as predicted by the SEEV model. However, surprisingly, the direction of the effect was opposite of what was predicted. Thus, these results do not support the second hypothesis. Specifically, the low-bandwidth display attracted more visual attention than the high-bandwidth display in the current study. Yet, these results are not necessarily inconsistent with the SEEV model because the model does not dictate the expectancy value of each AOI. Rather, when applying the SEEV model to a display, the user of the model must assign an expectancy value to each AOI in the display (relative to all other AOIs) *a priori* based on general guidance provided with the SEEV model. This associated guidance describes how specific properties of an AOI can influence the four parameters of the model (Wickens, 2015). For example, the SEEV guidance indicates that an AOI with higher information bandwidth should typically be assigned a higher expectancy value, and based on the SEEV equation (Equation 2), higher expectancy always leads to increased attention allocation (assuming the value parameter in the model is not zero).

In the current study, rather than comparing different AOIs in a single display, the relative attentional attraction of a single AOI with two different information bandwidth settings was examined. The prediction regarding the relationship between information bandwidth and expectancy was based on the general expectancy guidance provided with the SEEV model. However, even though this prediction is recommended in most cases, the expectancy guidance also indicates that expectancy can be influenced by context (Wickens et al., 2022). This can occur in situations where a contextual cue (e.g., an alert in a different AOI) suggests that information has become available in an AOI, momentarily prioritizing the cued AOI. The guidance suggests that this could potentially lead to increased probability of sampling a low-bandwidth AOI. Unfortunately, further specific guidance is not provided, and supporting evidence is minimal (Wickens et al., 2022). Thus, the SEEV model accounts for the results observed in the current study, but the guidance was insufficient to establish an accurate hypothesis.

General guidance associated with the SEEV model defines expectancy as the operator's expectation of obtaining information from an AOI, with contributing factors divided into two categories: information bandwidth and context (Wickens et al., 2022). As previously described, the bandwidth-based guidance indicates that operators will tend to sample high-bandwidth AOIs more often than low-bandwidth AOIs. This pattern is based on strong evidence from a series of classic studies conducted by Senders and colleagues (Carbonell et al., 1968; Senders, 1964, 1983). For example, Senders (1964) asked participants to monitor a group of gauges with moving pointers and identify when any individual gauge reading exceeded the nominal range. Results from this study indicated that the gauges were sampled proportionally to their information bandwidth. Some of these studies have since been replicated, and the original findings have been confirmed (e.g., Eisma et al., 2020). Typically, when the SEEV model is applied in a study, information bandwidth is the only factor discussed when referencing expectancy (e.g., Horrey et al., 2006; Wickens, 2015). Similarly, in the current study, the expectancy prediction was based solely on information bandwidth as there was no specific evidence to suggest a different effect given the display and task configurations included in the study.

Based on RT and PDT results in the current study, it appears that participants who utilized the supplemental data were triggered to view the information in the SDD AOI after being alerted of an autopilot failure from the resource management task AOI. After transitioning their gaze to the SDD AOI, participants would then wait for the relevant information to appear (i.e., one of the temperature values in the SDD graph rising above the threshold) before transitioning back to the resource management task AOI to resolve the autopilot failure. Because the relevant SDD information was slower to appear in the Low Bandwidth condition, participants spent more time waiting for the information while viewing the SDD AOI in the Low Bandwidth compared to the High Bandwidth condition. This description of contextual cueing aligns with observations from the current study, but the described behavior is only assumed. Further exploration is required to confirm these assumptions by analyzing individuals' eye movement data relative to the specific timing of failure events.

VALUE AND EXPECTANCY IN THE SEEV MODEL

The relationship between the expectancy and value parameters of the SEEV model in the literature has not been consistent. The current study was designed to test whether the relationship between these parameters is additive or interactive. The latter was predicted based on the most recent publication of the SEEV model (Wickens et al., 2022). In the current study, there was substantial variability in task relevance (i.e., value parameter) of the Supplemental Data Display information within each mental model group as observed through performance measures (see Figure 15 and Figure 16). This variability was presumably present because the mental model manipulation was intentionally designed to avoid directly influencing task relevance (i.e., participants were required to infer task relevance based on their mental models of the system). This variability limits the utility of the mental model groups as task value groups for the purpose of understanding the relationship between the expectancy and value parameters in the SEEV model. Therefore, to effectively evaluate the relationship between value and expectancy, more stringent performance-based criteria were applied to the groups to ensure that each group more closely reflected differences in task value.

The results indicated a strong interaction between task value group and information bandwidth (Figure 17), which supports the third hypothesis that the relationship between the expectancy and value parameters in the SEEV model is interactive. In other words, the effect of changes in expectancy (i.e., information bandwidth) is negligible when the information that is displayed has no value to the participant. Because this hypothesis also predicted the direction of the bandwidth effect, which was opposite of the observed effect (as discussed in the previous section), the third hypothesis is only partially supported. The interaction is generally consistent with Horrey and colleagues' (2006) applied driving research that examined information bandwidth and task priority. Although, the presence of the interaction in their study was dependent on how information bandwidth was operationalized. The results of the current study provide evidence for an additional implementation of information bandwidth in a more controlled environment.

In the current study, the training intervention provided each group with slightly different information about how the autopilot system works, which led to differential patterns of participants' attention allocation. The effect of this training manipulation on mental models is consistent with previous mental model research (e.g., Weis & Wiese, 2022; Politowicz et al., 2022). Furthermore, the observed influence of mental models on performance (e.g., Sarter et al., 2007) and attention allocation (e.g., Koh et al., 2011) is consistent with previous research. However, it is important to understand how the training manipulation led to participants being categorized as having mixed SDD use (Figure 15). Presumably, these thirteen participants modified their mental models of the automation during exposure to the system (i.e., during the experimental trials) to the extent that the value of the SDD AOI increased substantially and influenced their visual attention allocation strategy. This aligns with Norman's (1983; 1986) theoretical model in which the user's model (i.e., mental model) continues to evolve with exposure to the system image (e.g., user interfaces). This is also consistent with performance findings of Kieras and Bovair (1984), who observed that providing process-based information will not always be of value (i.e., will not affect performance) and that some users may be able to infer process-based information without explicit instruction. The current study extends these results from performance to visual attention allocation.

THEORETICAL AND PRACTICAL IMPLICATIONS

From a theoretical perspective, the current study produced several interesting findings that impact the understanding and empirical basis of the SEEV model. First, results of this study provide strong empirical evidence to support the interactive relationship of expectancy and value in the SEEV model rather than the additive relationship. This relationship has previously been described as both additive (Wickens, 2015, 2021; Wickens & McCarley, 2008) and interactive (Wickens et al., 2001, 2008, 2022). However, the evidence that led to this discrepancy was primarily based on model fitting of experimental data (Wickens et al., 2008). The current study, on the other hand, employed a direct manipulation of the two parameters, which provides more conclusive evidence through causal inference. Furthermore, expectancy and value are generally considered to be the most important parameters in the SEEV model because optimal scanning is defined by these two parameters alone (e.g., Grundgeiger et al., 2020; Wickens, 2015). Thus, it is particularly important to understand the relationship between these two parameters.

Second, results from the current study indicate that the value parameter of the SEEV model is sensitive to differences in mental models of an automated system used during a task, which adds to previous findings that have primarily shown the value parameter is sensitive to differences in expertise (i.e., novice, expert; e.g., Koh et al., 2011) and differences in mental models of the tasks (i.e., task priorities; e.g., Horrey et al., 2006). This finding deepens the overall understanding of contributing factors that guide visual attention allocation during supervisory control tasks.

Lastly, results of the current study indicate that SEEV model guidance regarding expectancy could be improved with additional empirical evidence (specifically, when and how expectancy may deviate from typical behavior). The current study demonstrated that assigning expectancy values to AOIs requires an understanding of how specific information will be used by the operator to complete the task. Simply relying on information bandwidth (i.e., information refresh rate) without this additional consideration could lead to incorrect expectancy determinations and thus diminish the overall utility of the SEEV model for predicting attention allocation. Furthermore, this finding also introduces the possibility that previous issues of observed data not fitting with SEEV model predictions (e.g., Wickens et al., 2008, which previously led to the discrepancy regarding the expectancy and value parameter relationship) could be due to incorrect assignment of expectancy values rather than problems with the SEEV model equation. However, this potential explanation is speculative, and additional research is required to confirm it.

From a practical perspective, the results of the current study indicate that providing operators with minimal additional process-based training can significantly alter how they allocate attention towards a display. This effect was observed with the addition of only four process-based training slides in a 79-slide training package. Furthermore, this minimal additional training can significantly improve operators’ task performance as a result of their altered visual attention allocation. Beyond these training implications, this finding also implies that eye tracking data could potentially be used as a method for detecting inaccuracies in users’ mental models of complex systems, which could then be remedied with training interventions. More broadly, improvements to the general understanding of visual attention allocation in supervisory control tasks is critical for technology development across many domains, including surface transportation and Advanced Air Mobility. Real-time eye tracking is becoming increasingly common in these operational environments (e.g., Liang et al., 2007), and understanding the underlying meaning of these data is essential, particularly when the automated or autonomous system utilizing the data is applying it to safety-critical decisions.

LIMITATIONS

The current study had several limitations. First, the study employed one specific task and system configuration. Thus, it is possible that the observed effects in this study are context dependent, such that a different task or system configuration could lead to different results. The discussion of the results attempted to account for this, but it should still be considered more broadly with respect to the findings. Second, the tasks employed in this study were performed by inexperienced participants after only a brief training session. Hence, the tasks performed in this study do not directly reflect tasks performed by actual remote aircraft operators. Therefore, the results should be validated in applied settings. Lastly, this study evaluated mental models of a novel system, which may not generalize to operators performing tasks with familiar systems. Again, the results should be validated in applied settings to confirm that they generalize beyond the laboratory setting.

FUTURE RESEARCH

Future research should evaluate the same paradigm with different task and system configurations to understand how the observed results generalize to other tasks. Additionally, factors such as trust may influence sampling behavior (e.g., Sato et al., 2023), so future research should investigate the influence of trust and other contributory factors on the relationship between mental models and sampling behavior. Furthermore, additional studies should be conducted to confirm the observed results in an applied environment with operators who have previous experience with the automated system. Lastly, additional research should be conducted to establish a stronger understanding of the various factors that influence expectancy.

CHAPTER V

CONCLUSIONS

This study sought to investigate the role of expectancy and mental models of automated systems in visual attention allocation during supervisory control tasks. Results of the study indicate that the value parameter of the SEEV model is sensitive to differences in mental models of an automated system used during a task. The study provides evidence in support of a multiplicative relationship between expectancy and value in the SEEV model. Lastly, the effect of expectancy due to contextual cues is not well defined, so additional research is required to understand the factors that can influence the change in visual attention allocation as a result of changes in information bandwidth.

REFERENCES

Aubuchon, V. V., Hashemi, K. E., Shively, R. J., & Wishart, J. M. (2022). Multi-vehicle (m:N) operations in the NAS - NASA's research plans. *AIAA AVIATION 2022 Forum*. AIAA AVIATION 2022 Forum, Chicago, IL & Virtual. https://doi.org/10.2514/6.2022-3758

Brams, S., Ziv, G., Levin, O., Spitz, J., Wagemans, J., Williams, A. M., & Helsen, W. F. (2019). The relationship between gaze behavior, expertise, and performance: A systematic review. *Psychological Bulletin*, *145*(10), 980–1027. https://doi.org/10.1037/bul0000207

Cavanagh, P., & Alvarez, G. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, *9*(7), 349–354. https://doi.org/10.1016/j.tics.2005.05.009

Carbonell, J., Ward, J., & Senders, J. (1968). A Queueing Model of Visual Sampling Experimental Validation. *IEEE Transactions on Man Machine Systems*, *9*(3), 82–87. https://doi.org/10.1109/TMMS.1968.300041

Cegarra, J., Valéry, B., Avril, E., Calmettes, C., & Navarro, J. (2020). OpenMATB: A Multi-Attribute Task Battery promoting task customization, software extensibility and experiment replicability. *Behavior Research Methods*, *52*(5), 1980–1990. https://doi.org/10.3758/s13428-020-01364-w

Comstock, J. R., & Arnegard, R. J. (1992). The Multi-Attribute Task Battery for human operator workload and strategic behavior research. *NASA NTRS*. https://ntrs.nasa.gov/citations/19920007912

Craik, K. J. W. (1943). *The nature of explanation*. Cambridge University Press.

de Winter, J. C. F. (2019). Pitfalls of automation: A faulty narrative?: Commentary on Hancock (2019) Some pitfalls in the promises of automated and autonomous vehicles. *Ergonomics*, *62*(4), 505–508. https://doi.org/10.1080/00140139.2019.1563334

Dearholt, D. W., & Schvaneveldt, R. W. (1990). Properties of Pathfinder networks. In R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization.* (pp. 1–30). Ablex Publishing.

Eisma, Y. B., Hancock, P. A., & de Winter, J. C. F. (2020). On Senders's models of visual sampling behavior. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *65*(5), 723–736. https://doi.org/10.1177/0018720820959956

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *59*(1), 5–27. https://doi.org/10.1177/0018720816681350

Furlough, C. S., & Gillan, D. J. (2018). Mental models: Structural differences and the role of experience. *Journal of Cognitive Engineering and Decision Making*, *12*(4), 269–287. https://doi.org/10.1177/1555343418773236

Goldsmith, T. E., & Davenport, D. M. (1990). Assessing structural similarity of graphs. In R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization.* (pp. 75–87). Ablex Publishing.

Goodrich, M. A., & Boer, E. R. (2003). Model-based human-centered task automation: A case study in ACC system design. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *33*(3), 325–336. https://doi.org/10.1109/TSMCA.2003.817040

Grundgeiger, T., Wurmb, T., & Happel, O. (2020). Statistical modeling of visual attention of junior and senior anesthesiologists during the induction of general anesthesia in real and simulated cases. *IEEE Transactions on Human-Machine Systems*, *50*(4), 317–326. https://doi.org/10.1109/THMS.2020.2983817

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). *Metrics for explainable AI: Challenges and prospects*. https://doi.org/10.48550/ARXIV.1812.04608

Horrey, W. J., Wickens, C. D., & Consalus, K. P. (2006). Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied*, *12*(2), 67–78. https://doi.org/10.1037/1076-898X.12.2.67

Interlink. (2017a). *JPathfinder user manual*. http://interlinkinc.net/

Interlink. (2017b). *JTarget* [Computer software]. Interlink. http://interlinkinc.net/

Interlink. (2021). *Pathfinder* (Version 9.0) [Computer software]. Interlink. http://interlinkinc.net/

Karpinsky, N. D., Chancey, E. T., Palmer, D. B., & Yamani, Y. (2018). Automation trust and attention allocation in multitasking workspace. *Applied Ergonomics*, *70*, 194–201. https://doi.org/10.1016/j.apergo.2018.03.008

Kieras, D. E., & Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science*, *8*(3), 255–273. https://doi.org/10.1207/s15516709cog0803_3

Koh, R. Y. I., Park, T., Wickens, C. D., Ong, L. T., & Chia, S. N. (2011). Differences in attentional strategies by novice and experienced operating theatre scrub nurses. *Journal of Experimental Psychology: Applied*, *17*(3), 233–246. https://doi.org/10.1037/a0025171

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. https://doi.org/10.1518/hfes.46.1.50_30392

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*(10), 1243–1270. https://doi.org/10.1080/00140139208967392

Liang, Y., Reyes, M. L., & Lee, J. D. (2007). Real-time detection of driver cognitive distraction using support vector machines. *IEEE Transactions on Intelligent Transportation Systems*, *8*(2), 340–350. https://doi.org/10.1109/TITS.2007.895298

Lim, B.-C., & Klein, K. J. (2006). Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior*, *27*(4), 403–418. https://doi.org/10.1002/job.387

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed). Lawrence Erlbaum Associates.

Moray, N. (1999). Mental models in theory and practice. In Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application. (pp. 223–258). The MIT Press.

Mutzenich, C., Durant, S., Helman, S., & Dalton, P. (2021). Updating our understanding of situation awareness in relation to remote operators of autonomous vehicles. *Cognitive Research: Principles and Implications*, *6*(1), 9. https://doi.org/10.1186/s41235-021-00271-8

National Academies of Sciences, Engineering, and Medicine. (2020). *Advanced aerial mobility: A national blueprint*. National Academies Press. https://doi.org/10.17226/25646

Norman, D. A. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 7–14). Erlbaum.

Norman, D. A. (1986). Cognitive engineering. In D. A. Norman & S. W. Draper (Eds.), *User centered system design: New perspectives on human-computer interaction* (pp. 31–61). Erlbaum.

Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *56*(3), 476–488. https://doi.org/10.1177/0018720813501549

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Patterson, M. D., Isaacson, D. R., Mendonca, N. L., Neogi, N. A., Goodrich, K. H., Metcalfe, M., Bastedo, B., Metts, C., Hill, B. P., DeCarme, D., Griffin, C., & Wiggins, S. (2021). An initial concept for intermediate-state, passenger-carrying Urban Air Mobility operations. *AIAA Scitech 2021 Forum*. AIAA Scitech 2021 Forum, VIRTUAL EVENT. https://doi.org/10.2514/6.2021-1626

Politowicz, M. S., Sato, T., Chancey, E. T., & Yamani, Y. (2022). Pathfinder networks for measuring operator mental model structure with a simple autopilot system. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 883–887. https://doi.org/10.1177/1071181322661510

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174. https://doi.org/10.1037/0096-3445.109.2.160

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, *100*(3), 349–363. https://doi.org/10.1037/0033-2909.100.3.349

Rowe, A. L., & Cooke, N. J. (1995). Measuring mental models: Choosing the right tools for the job. *Human Resource Development Quarterly*, *6*(3), 243–255. https://doi.org/10.1002/hrdq.3920060303

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, *115*(1), 101–130. https://doi.org/10.1037/0033-295X.115.1.101

Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2007). Pilots' monitoring strategies and performance on automated flight decks: An empirical study combining behavioral and eye-tracking data. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(3), 347–357. https://doi.org/10.1518/001872007X196685

Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 5–19. https://doi.org/10.1518/001872095779049516

Sato, T., Inman, J., Politowicz, M. S., Chancey, E. T., & Yamani, Y. (2023). A meta-analytic approach to investigating the relationship between human-automation trust and attention allocation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. https://doi.org/10.1177/21695067231194333

Sato, T., Yamani, Y., Liechty, M., & Chancey, E. T. (2020). Automation trust increases under high-workload multitasking scenarios involving risk. *Cognition, Technology & Work*, *22*(2), 399–407. https://doi.org/10.1007/s10111-019-00580-5

Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). Network structures in proximity data. In *Psychology of Learning and Motivation* (Vol. 24, pp. 249–284). Elsevier. https://doi.org/10.1016/S0079-7421(08)60539-3

Schvaneveldt, R. W., Durso, F. T., Goldsmith, T. E., Breen, T. J., Cooke, N. M., Tucker, R. G., & De Maio, J. C. (1985). Measuring the structure of expertise. *International Journal of Man-Machine Studies*, *23*(6), 699–728. https://doi.org/10.1016/S0020-7373(85)80064-X

Senders, J. W. (1964). The human operator as a monitor and controller of multidegree of freedom systems. *IEEE Transactions on Human Factors in Electronics*, *HFE-5*(1), 2–5. https://doi.org/10.1109/THFE.1964.231647

Senders, J. W. (1983). *Visual scanning processes* [Doctoral Thesis, Tilburg University].

Tossell, C. C., Schvaneveldt, R. W., & Branaghan, R. J. (2010). Targeting knowledge structures: A new method to elicit the relatedness of concepts. *Cognitive Technology*, *15*(2), 11–19.

Underwood, G., Chapman, P., Bowden, K., & Crundall, D. (2002). Visual search while driving: Skill and awareness during inspection of the scene. *Transportation Research Part F: Traffic Psychology and Behaviour*, *5*(2), 87–97. https://doi.org/10.1016/S1369-8478(02)00008-6

Weis, P. P., & Wiese, E. (2022). Know your cognitive environment! Mental models as crucial determinant of offloading preferences. *Human Factors: The Journal of the Human*

Factors and Ergonomics Society, *64*(3), 499–513. https://doi.org/10.1177/0018720820956861

Wickens, C. (2021). Attention: Theory, principles, models and applications. *International Journal of Human–Computer Interaction*, *37*(5), 403–417. https://doi.org/10.1080/10447318.2021.1874741

Wickens, C. D. (2015). Noticing events in the visual workplace: The SEEV and NSEEV models. In R. R. Hoffman, P. A. Hancock, M. W. Scerbo, R. Parasuraman, & J. L. Szalma (Eds.), *The Cambridge Handbook of Applied Perception Research* (pp. 749–768). Cambridge University Press. https://doi.org/10.1017/CBO9780511973017.046

Wickens, C. D., Gutzwiller, R. S., & McCarley, J. S. (2022). *Applied attention theory* (2nd ed.). CRC Press. https://doi.org/10.1201/9781003081579

Wickens, C. D., Helleberg, J., Goh, J., Xu, X., & Horrey, W. J. (2001). Pilot task management: Testing an attentional expected value model of visual scanning. *Savoy, IL, UIUC Institute of Aviation Technical Report*.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering psychology and human performance* (Fourth edition). Pearson.

Wickens, C. D., & McCarley, J. S. (2008). *Applied attention theory*. CRC Press.

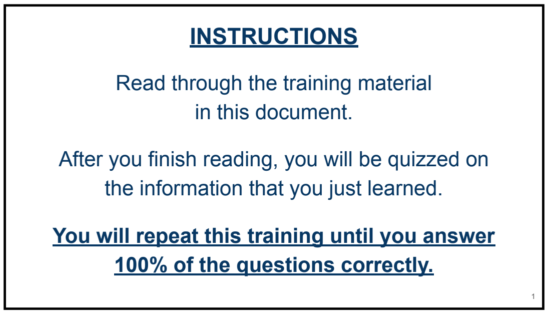
Wickens, C. D., McCarley, J. S., Alexander, A. L., Thomas, L. C., Ambinder, M., & Zheng, S. (2008). Attention-situation awareness (A-SA) model of pilot error. *Human Performance Modeling in Aviation*, 213–239.

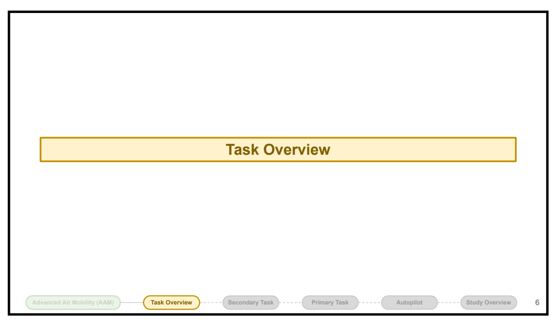
Ziv, G. (2016). Gaze behavior and visual attention: A review of eye tracking studies in aviation. *The International Journal of Aviation Psychology*, *26*(3–4), 75–104. https://doi.org/10.1080/10508414.2017.1313096

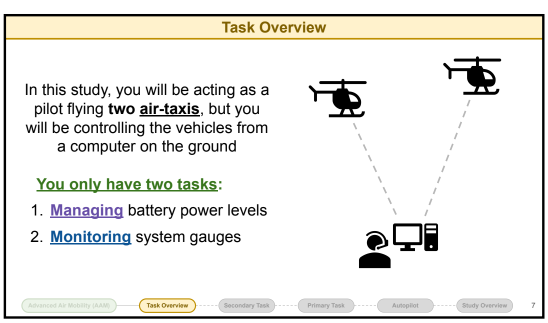
APPENDIX A

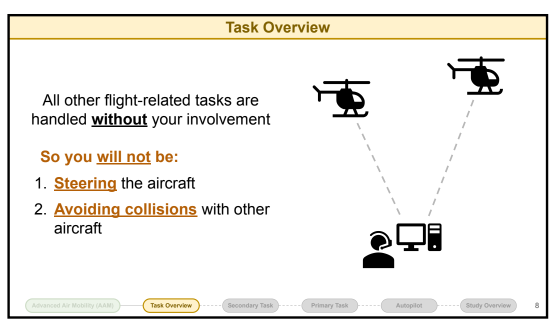
TRAINING SLIDES: BOTH MENTAL MODEL GROUPS

Note: *Slides 69-72 are specific to each Mental Model group. Both sets are included on page 88.*



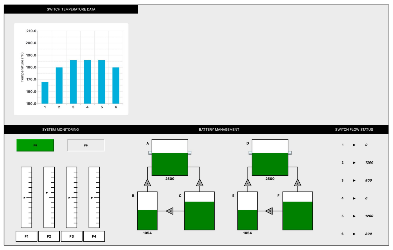






Task Overview

This is the display that you will use to complete your tasks

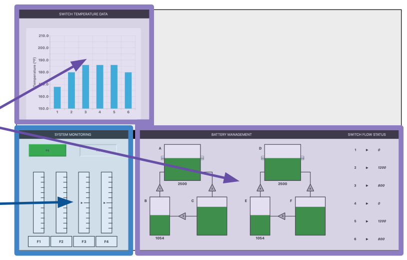


Task Overview

You will have two continuous tasks

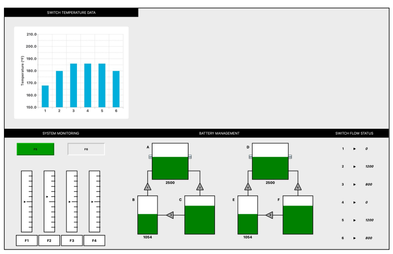
Primary task: **Battery Management**

Secondary task: **System Monitoring**



Task Overview

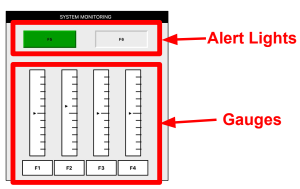
You will now learn how to perform each task…



SECONDARY TASK: System Monitoring Task

SECONDARY TASK: System Monitoring Task

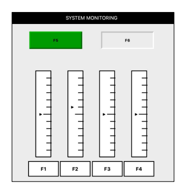
This task requires you to monitor **4 gauges** and **2 alert lights**



SECONDARY TASK: System Monitoring Task

There is **no automation** to assist you with this task

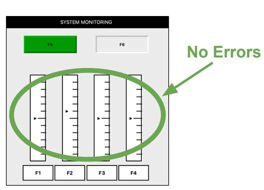
You simply have to monitor the **gauges** and **alert lights** for **errors**, and then press a button to fix each **error**



SECONDARY TASK: System Monitoring Task

The small black triangle in each **gauge** needs to remain **near the center**

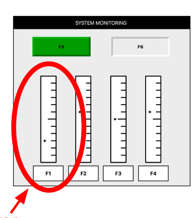
The triangle will bounce around, which is okay as long as it stays close to the center

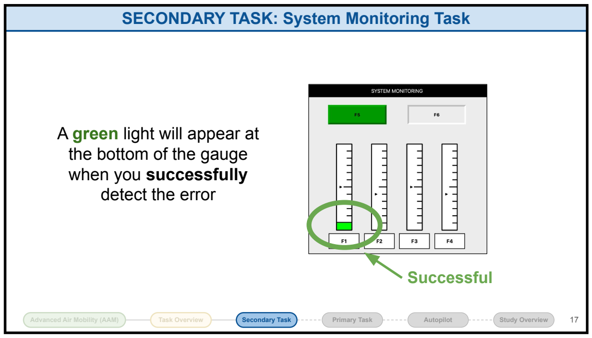


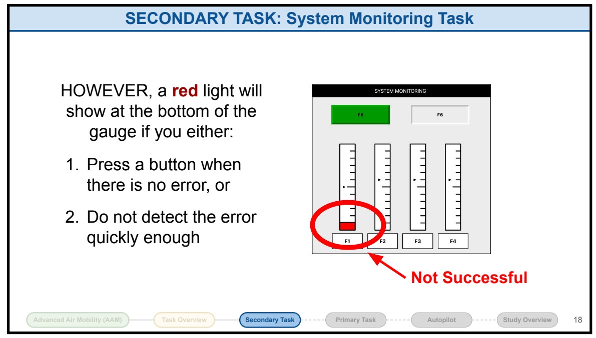
SECONDARY TASK: System Monitoring Task

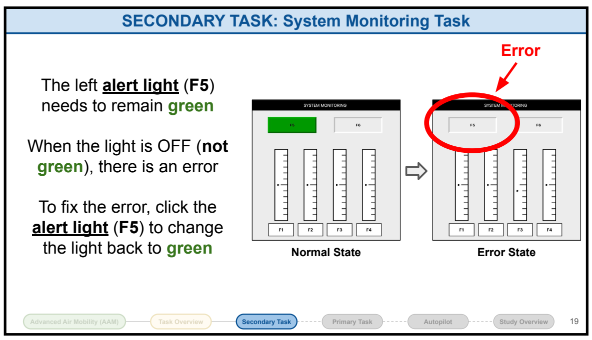
An **error** occurs when the triangle in a **gauge** moves **near the top or bottom**

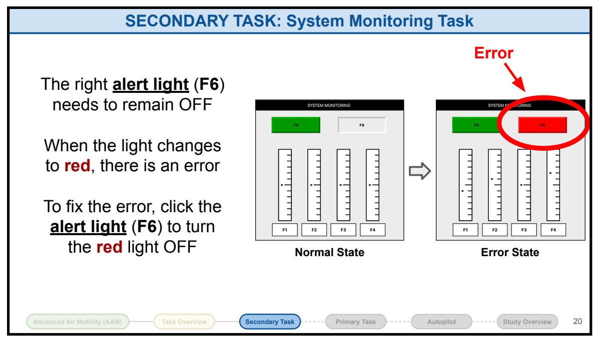
To fix the error, use the mouse to click the **button** (F1, F2, F3, or F4) below the **gauge** that is experiencing the error

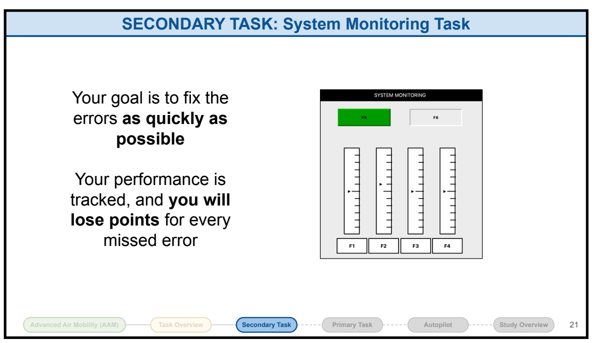


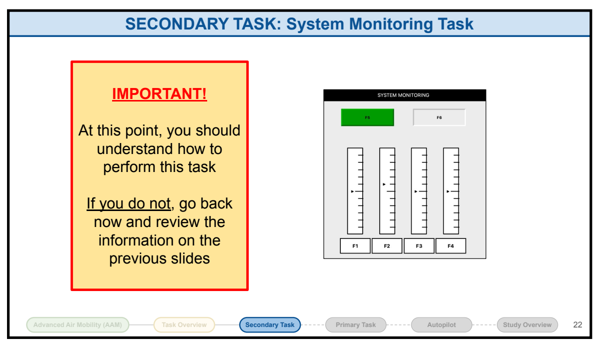

SECONDARY TASK: System Monitoring Task
A green light will appear at the bottom of the gauge when you successfully detect the error
SYSTEM MONITORING
F5
F6
F1
F2
F3
F4
Successful
Advanced Air Mobility (AAM)
Task Overview
Secondary Task
Primary Task
Autopilot
Study Overview
17

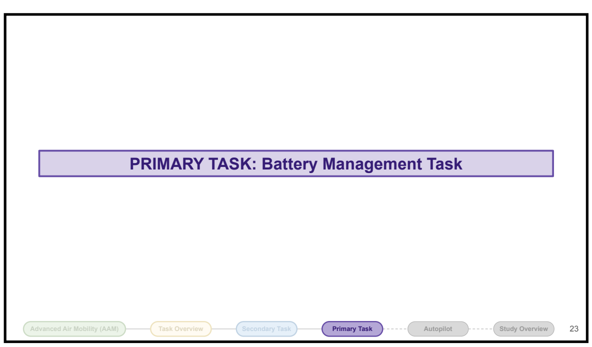

SECONDARY TASK: System Monitoring Task
HOWEVER, a red light will show at the bottom of the gauge if you either:
1. Press a button when there is no error, or
2. Do not detect the error quickly enough
SYSTEM MONITORING
F5
F6
F1
F2
F3
F4
Not Successful
Advanced Air Mobility (AAM)
Task Overview
Secondary Task
Primary Task
Autopilot
Study Overview
18

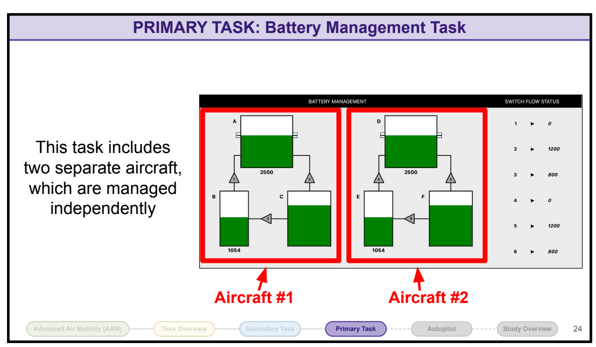

SECONDARY TASK: System Monitoring Task
The left alert light (F5) needs to remain green
When the light is OFF (not green), there is an error
To fix the error, click the alert light (F5) to change the light back to green
Error
Normal State
Error State
Advanced Air Mobility (AAM)
Task Overview
Secondary Task
Primary Task
Autopilot
Study Overview
19


SECONDARY TASK: System Monitoring Task
The right alert light (F6) needs to remain OFF
When the light changes to red, there is an error
To fix the error, click the alert light (F6) to turn the red light OFF
Error
Normal State
Error State
Advanced Air Mobility (AAM)
Task Overview
Secondary Task
Primary Task
Autopilot
Study Overview
20


SECONDARY TASK: System Monitoring Task
Your goal is to fix the errors as quickly as possible
Your performance is tracked, and you will lose points for every missed error
SYSTEM MONITORING
Advanced Air Mobility (AAM)
Task Overview
Secondary Task
Primary Task
Autopilot
Study Overview
21

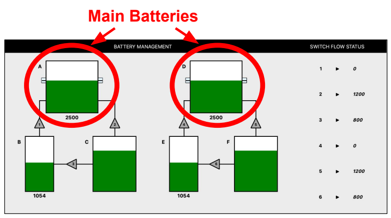

SECONDARY TASK: System Monitoring Task
IMPORTANT!
At this point, you should understand how to perform this task
If you do not, go back now and review the information on the previous slides
SYSTEM MONITORING
Advanced Air Mobility (AAM)
Task Overview
Secondary Task
Primary Task
Autopilot
Study Overview
22


PRIMARY TASK: Battery Management Task
Advanced Air Mobility (AAM)
Task Overview
Secondary Task
Primary Task
Autopilot
Study Overview
23


PRIMARY TASK: Battery Management Task
This task includes two separate aircraft, which are managed independently
BATTERY MANAGEMENT
SWITCH FLOW STATUS
Aircraft #1
Aircraft #2
Advanced Air Mobility (AAM)
Task Overview
Secondary Task
Primary Task
Autopilot
Study Overview
24

PRIMARY TASK: Battery Management Task

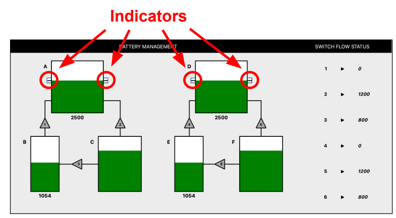
For this task, you must maintain the appropriate battery energy level for the **main battery** on each aircraft



PRIMARY TASK: Battery Management Task

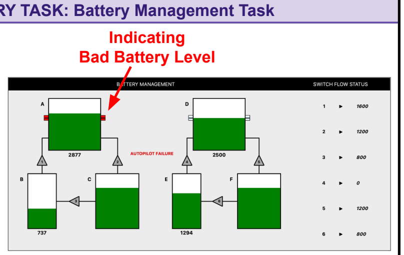
The **green color** inside the **battery** shows the current power level (this will move up & down)

The power level must remain within the top & bottom of the indicators on the sides of the **battery**



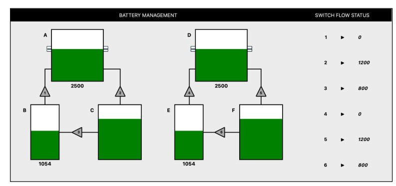
PRIMARY TASK: Battery Management Task

If the **battery** power level goes above or below the indicator, the indicator will turn **red** (as shown in the image on the right)



PRIMARY TASK: Battery Management Task

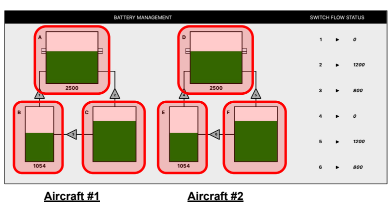
Each aircraft consists of **3 batteries** and **3 switches**



PRIMARY TASK: Battery Management Task

The **batteries** are highlighted in the image on the right and labeled as follows:

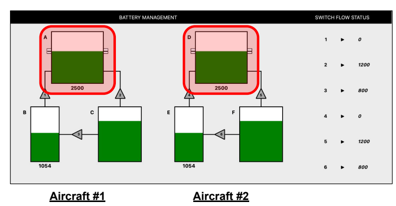
Aircraft #1: **A, B, C**
Aircraft #2: **D, E, F**



PRIMARY TASK: Battery Management Task

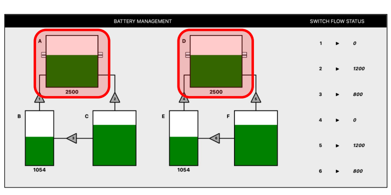
For Aircraft #1 (left), **Battery A** will slowly go down because it is being drained by the aircraft systems

It needs to be constantly replenished from the other two **batteries** (**B & C**)



PRIMARY TASK: Battery Management Task

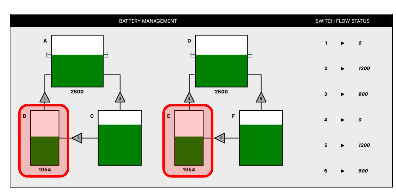
This is also true for Aircraft #2 (right), where **Battery D** is slowly drained and is constantly replenished by **Batteries E & F**



PRIMARY TASK: Battery Management Task

The power levels in **Batteries B & E** will only go up and down when power is being sent to or received from one of the other connected **batteries**

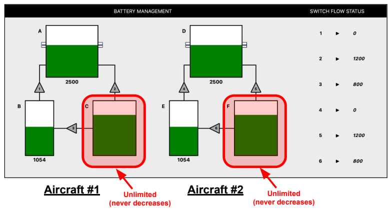
There is no target power level for **Batteries B & E** *(any level is allowed)*



PRIMARY TASK: Battery Management Task

The power levels in **Batteries C & F** will never move because they have unlimited power

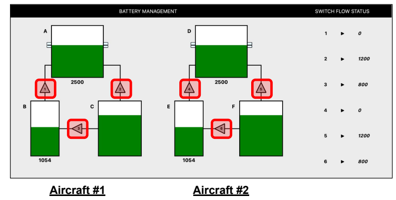
These two **batteries** are the main source of power for the other **batteries**



PRIMARY TASK: Battery Management Task

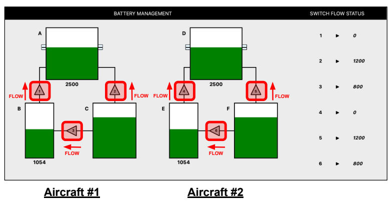
The **switches** are highlighted in the right image on the right and labeled as follows:

Aircraft #1: **1, 2, 3**
Aircraft #2: **4, 5, 6**



PRIMARY TASK: Battery Management Task

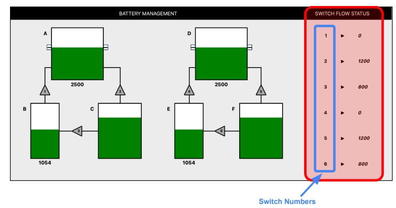
The **switches** turn ON / OFF the flow from one **battery** to another in the direction that the triangle is facing



PRIMARY TASK: Battery Management Task

The **Switch Flow Status** displays the current rate of electricity flowing through each **switch**

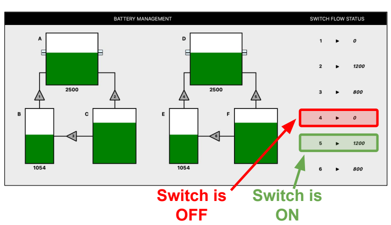
The numbers on the left in this display correspond to the numbers on each **switch**



PRIMARY TASK: Battery Management Task

You can tell when a **switch** is **ON** or **OFF** by looking at the values in the **Switch Flow Status** display

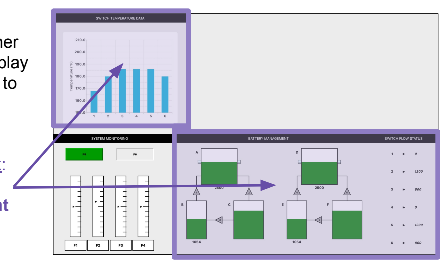
For example, in this image, **Switch 4** is **OFF** because the electricity flow is zero, & **Switch 5** is **ON** because it has a flow value



PRIMARY TASK: Battery Management Task

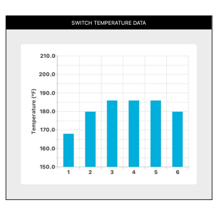
There is another view on the display that is related to this task

Primary task:
Battery Management



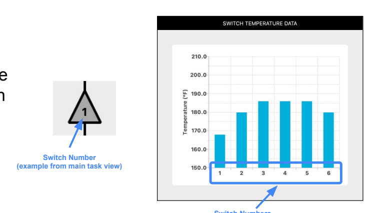
PRIMARY TASK: Battery Management Task

The **Switch Temperature Data** graph shows the current temperature of each **switch**



PRIMARY TASK: Battery Management Task

The numbers on the bottom of the graph correspond to the numbers on each **switch**

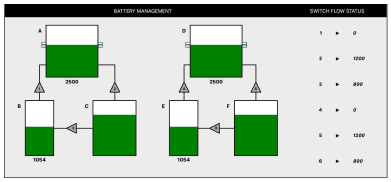


PRIMARY TASK: Battery Management Task

IMPORTANT!

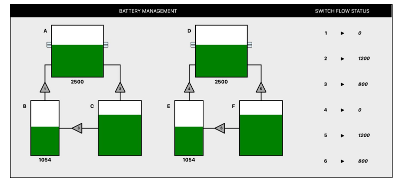
At this point, you should understand all of the components in the image on the right

If you do not, go back now and review the information on the previous slides



PRIMARY TASK: Battery Management Task

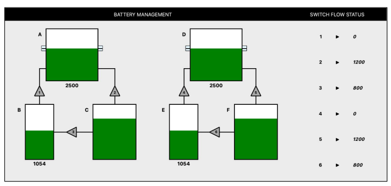
Again, your job is to maintain the battery level of the **main battery** for each aircraft (**A & D**)



PRIMARY TASK: Battery Management Task

Each vehicle is equipped with an automated system that will assist you with this task

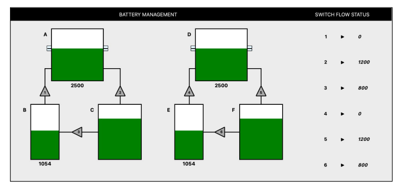
This system is called the **autopilot**



PRIMARY TASK: Battery Management Task

The **autopilot** is always on and assisting

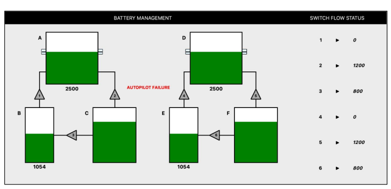
It automatically flips the **switches** ON / OFF to maintain the **main battery** level



PRIMARY TASK: Battery Management Task

But sometimes the **autopilot** will have a problem and **will fail** to maintain the correct **battery** level

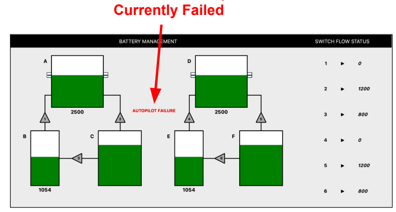
When this happens, an **autopilot failure** message will appear



PRIMARY TASK: Battery Management Task

Indicates that an Autopilot is Currently Failed

Note: The **autopilot failure** message does not specify *which* **autopilot** has failed (it could be Aircraft #1, Aircraft #2, or both)

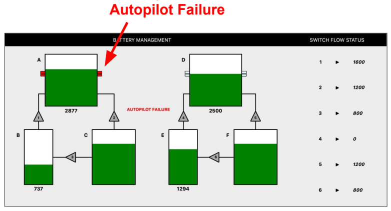


Aircraft #1 **Aircraft #2**

PRIMARY TASK: Battery Management Task

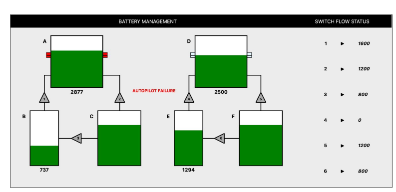
Bad Battery Level due to Autopilot Failure

After the **autopilot failure** message appears, the **autopilot will fail** to maintain the correct **battery** level



PRIMARY TASK: Battery Management Task

What do you do if the **autopilot** fails?

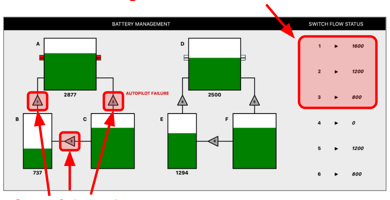


PRIMARY TASK: Battery Management Task

When the **autopilot** has **failed**, it means that one of the **switches** is stuck OFF

Even though the **Switch Flow Status** shows that all three switches are ON

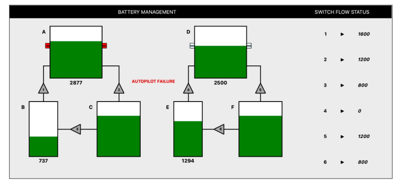
Showing All 3 Switch ON



One of these is really OFF

PRIMARY TASK: Battery Management Task

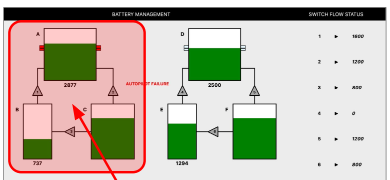
You have to figure out which **switch** is actually OFF and use the mouse to click that **switch** to reset and fix the **autopilot**



PRIMARY TASK: Battery Management Task

To figure out which **switch** is OFF, you can watch the flow of power between the **batteries** for a few seconds

For example, if **Battery B** is increasing, you can conclude that **Switch 1** is OFF



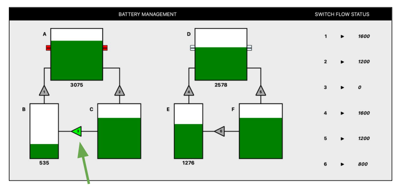
Watch how the power is moving between the batteries

PRIMARY TASK: Battery Management Task

Once you determine which **switch** is OFF, click that **switch**

If you clicked the *correct* **switch**, it will turn **green** to indicate that you were correct

The **autopilot** is now fixed!

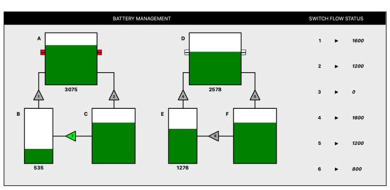


Indicates this switch was correctly clicked

PRIMARY TASK: Battery Management Task

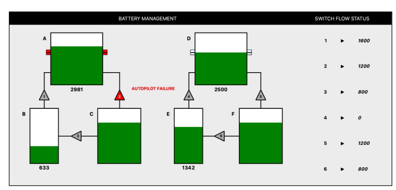
The **main battery** level will now start to move back towards the acceptable range

Note: The target indicators will stay **red** until the **main battery** level has had enough time to return to the acceptable range



PRIMARY TASK: Battery Management Task

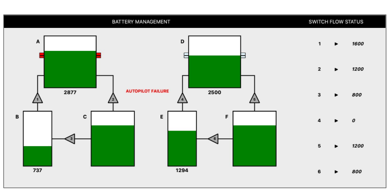
HOWEVER, if you click a **switch** that ***does not*** need to be clicked, it will turn **red**, and you will be locked out from clicking any other **switches** on that aircraft until the **switch** color changes back to gray (after **2 seconds**)



PRIMARY TASK: Battery Management Task

The **autopilot** will continue to malfunction until the correct **switch** is clicked!

So if you have not yet clicked the correct **switch**, you need to try another **switch**

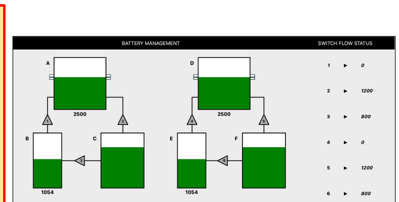


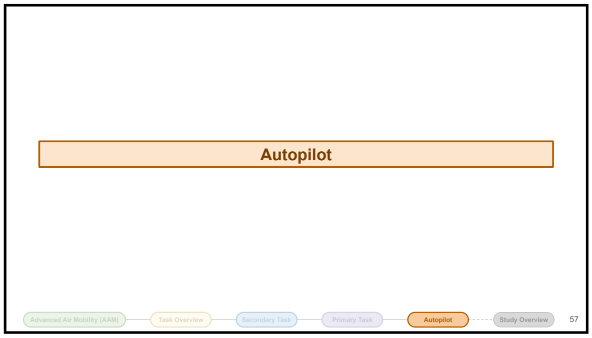
PRIMARY TASK: Battery Management Task

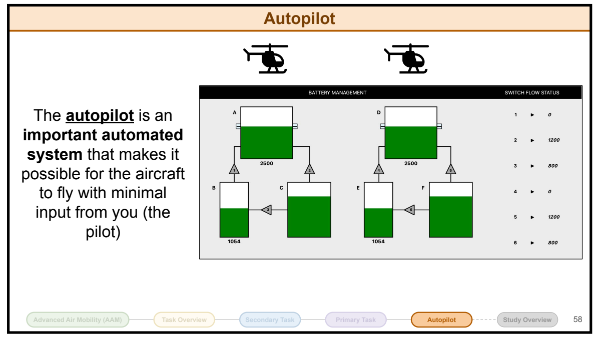
IMPORTANT!

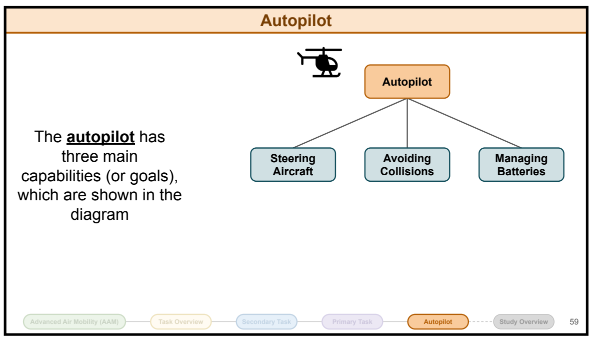
At this point, you should understand how to perform this task

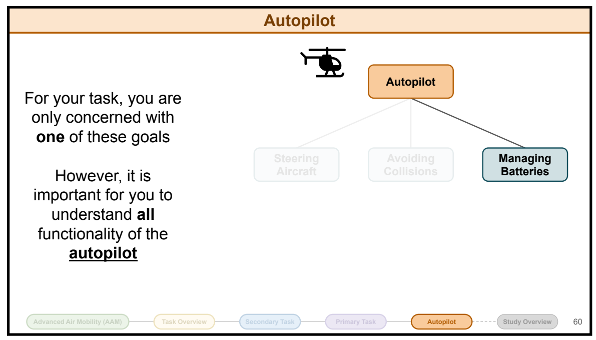
If you do not, go back now and review the information on the previous slides

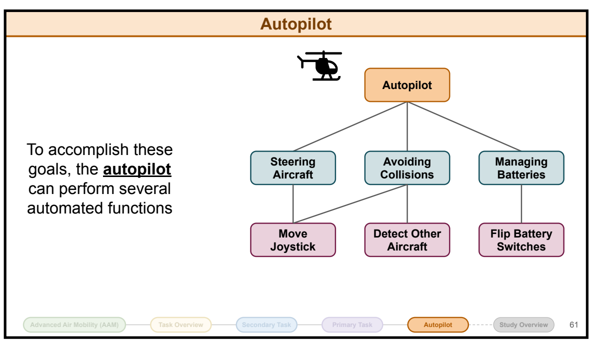


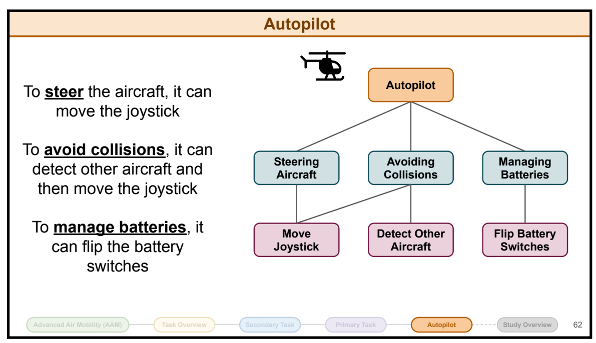

Autopilot
Advanced Air Mobility (AAM) Task Overview Secondary Task Primary Task Autopilot Study Overview
57

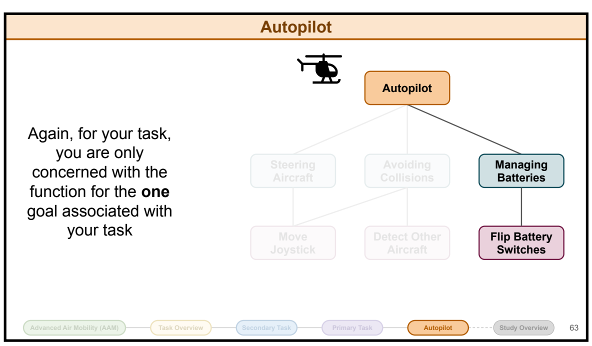

Autopilot
The **autopilot** is an **important automated system** that makes it possible for the aircraft to fly with minimal input from you (the pilot)
BATTERY MANAGEMENT
SWITCH FLOW STATUS
Advanced Air Mobility (AAM) Task Overview Secondary Task Primary Task Autopilot Study Overview
58

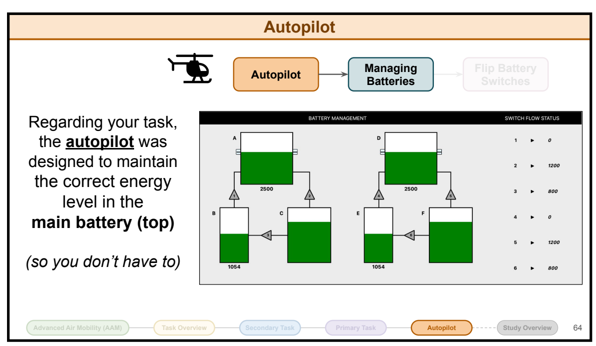

Autopilot
The **autopilot** has three main capabilities (or goals), which are shown in the diagram
Autopilot
Steering Aircraft
Avoiding Collisions
Managing Batteries
Advanced Air Mobility (AAM) Task Overview Secondary Task Primary Task Autopilot Study Overview
59


Autopilot
For your task, you are only concerned with **one** of these goals
However, it is important for you to understand **all** functionality of the **autopilot**
Autopilot
Steering Aircraft
Avoiding Collisions
Managing Batteries
Advanced Air Mobility (AAM) Task Overview Secondary Task Primary Task Autopilot Study Overview
60


Autopilot
To accomplish these goals, the **autopilot** can perform several automated functions
Autopilot
Steering Aircraft
Avoiding Collisions
Managing Batteries
Move Joystick
Detect Other Aircraft
Flip Battery Switches
Advanced Air Mobility (AAM) Task Overview Secondary Task Primary Task Autopilot Study Overview
61


Autopilot
To **steer** the aircraft, it can move the joystick
To **avoid collisions**, it can detect other aircraft and then move the joystick
To **manage batteries**, it can flip the battery switches
Autopilot
Steering Aircraft
Avoiding Collisions
Managing Batteries
Move Joystick
Detect Other Aircraft
Flip Battery Switches
Advanced Air Mobility (AAM) Task Overview Secondary Task Primary Task Autopilot Study Overview
62

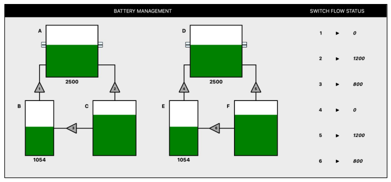

Autopilot
Again, for your task, you are only concerned with the function for the **one** goal associated with your task
Autopilot
Steering Aircraft
Avoiding Collisions
Managing Batteries
Move Joystick
Detect Other Aircraft
Flip Battery Switches
Advanced Air Mobility (AAM) Task Overview Secondary Task Primary Task Autopilot Study Overview
63


Autopilot
Autopilot
Managing Batteries
Flip Battery Switches
Regarding your task, the **autopilot** was designed to maintain the correct energy level in the **main battery (top)**
(so you don't have to)
BATTERY MANAGEMENT
SWITCH FLOW STATUS
Advanced Air Mobility (AAM) Task Overview Secondary Task Primary Task Autopilot Study Overview
64

Autopilot

Autopilot → Managing Batteries → Flip Battery Switches

To accomplish this, the **autopilot** automatically flips the **switches** ON / OFF as needed



Autopilot

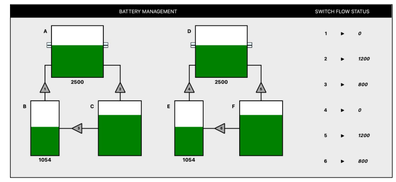
When the **autopilot** is successful:

The **switches** are flipped when needed

The **battery** levels remain in the range

The **main battery indicators** remain white

Autopilot Success



Autopilot

IMPORTANT!

The **autopilot** is reliable, but it could fail by **incorrectly setting the switches**

Autopilot

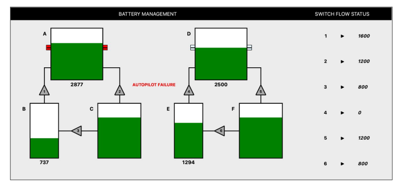
When the **autopilot** fails:

One of the **switches** gets stuck OFF (but the status shows ON)

The **battery** levels leave the target range

The **main battery indicators** turns red

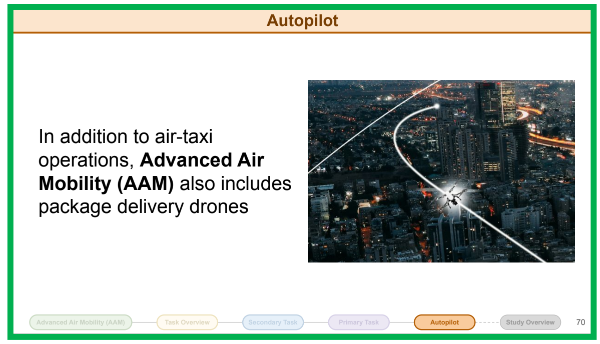
Autopilot Failure



<u>Note</u>: Slides 69-72 for each Mental Model group (Basic and Advanced) are depicted below.

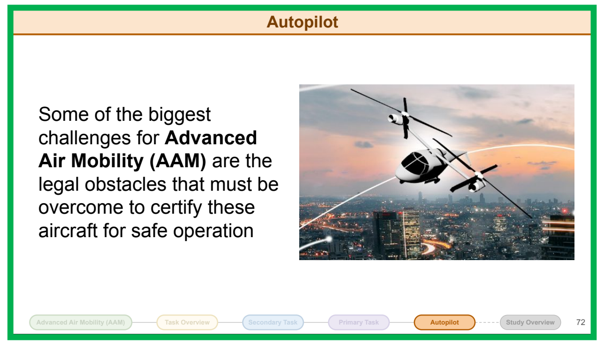
BASIC MENTAL MODEL GROUP ONLY

BASIC MENTAL MODEL GROUP ONLY

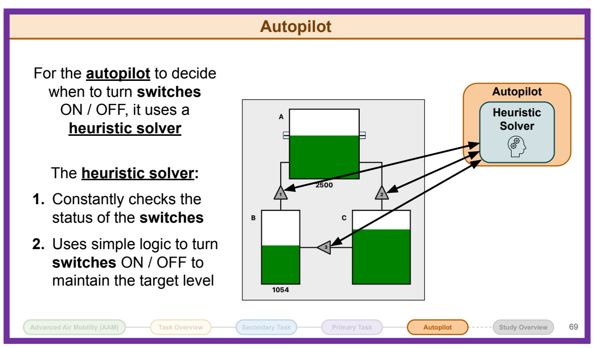


BASIC MENTAL MODEL GROUP ONLY

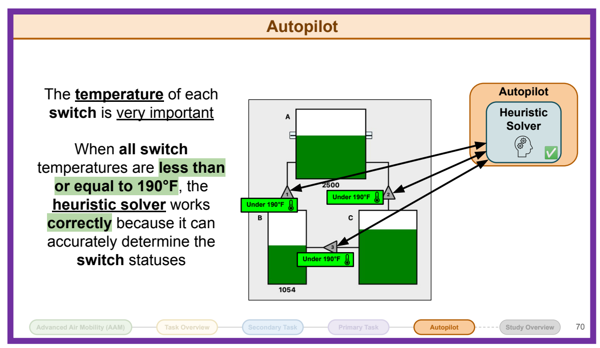
BASIC MENTAL MODEL GROUP ONLY



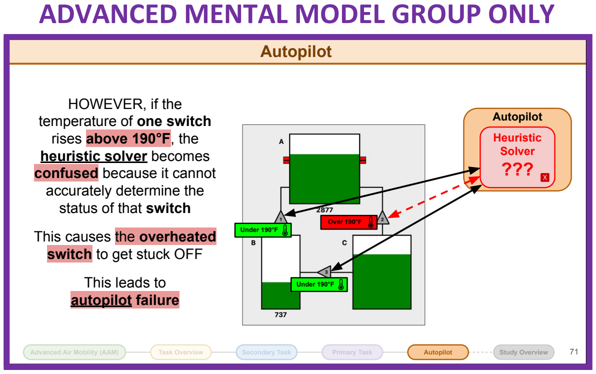
ADVANCED MENTAL MODEL GROUP ONLY



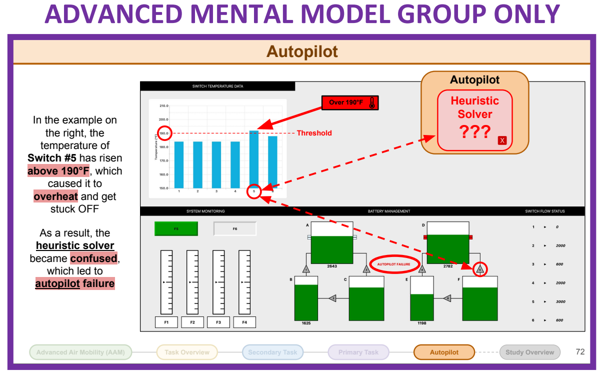
ADVANCED MENTAL MODEL GROUP ONLY



ADVANCED MENTAL MODEL GROUP ONLY



ADVANCED MENTAL MODEL GROUP ONLY



Study Overview

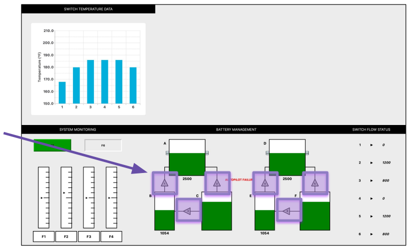
Study Overview

TASK RECAP

Battery Management:

Click these **6 switches** when there is an **autopilot failure**

This task should be your #1 priority



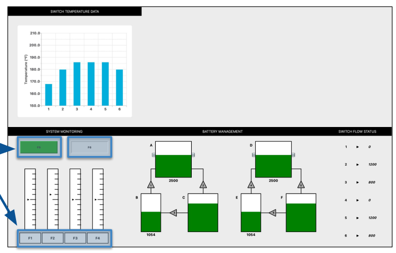
Study Overview

TASK RECAP

System Monitoring:

Click these **6 buttons** to fix errors as they appear

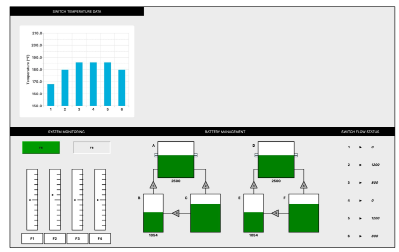
This task should be your #2 priority



Study Overview

Your goal is to perform the tasks as well as you possibly can

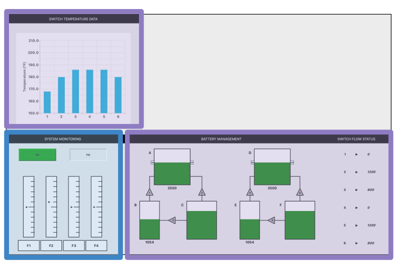
Every miss or incorrect guess **will count against you**



Study Overview

Important Note:

The two tasks function independently of one another, so the **System Monitoring** behavior is not connected to the **Battery Management**



Study Overview

Experiment Overview

- ➔ You will complete 4 separate flights (or trials)
- ➔ Each flight will last approximately 6 minutes
- ➔ Prioritize the primary task over the secondary task
- ➔ Perform as well as you can because your score will be recorded
- ➔ After the 4 trials, you will answer several questionnaires, which will end the experiment

This is the end of the training

You will now take a quiz covering the material that you just read.

You must score 100% on the quiz, or you will have to repeat this training.

When you are ready to continue:
Press ESC key (to exit full screen), then Click Here.

APPENDIX B

TRAINING ASSESSMENT QUESTIONNAIRES

Basic Mental Model Training Assessment

Note: Questions 8-12 are specific for the Basic Mental Model group.

Training

Answer the following questions based on the information that you learned from the training slides.

You must score 100% on the quiz, or you will have to repeat this training.

When you are done, review your score and wait for the researcher. **DO NOT RESTART THE QUIZ**.

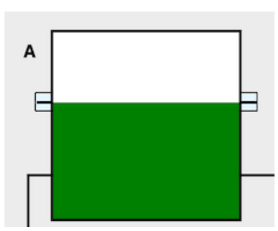
* Indicates required question

1. What are your two tasks in this study? * 1 point

Check all that apply.

- [ ] Steer the aircraft with a joystick
- [ ] Monitor the system gauges
- [ ] Manage the battery levels in each aircraft
- [ ] Avoid collisions with other aircraft

2. What is shown in this image? * 1 point



Mark only one oval.

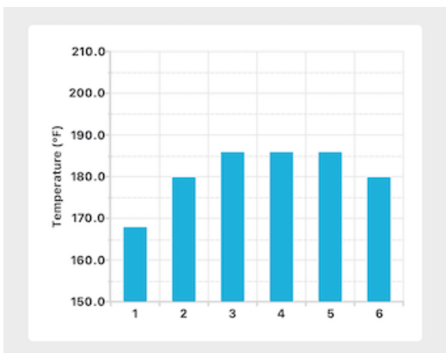
- ( ) A switch
- ( ) A battery

3. What is shown in this image? * 1 point

Mark only one oval.

- A switch
- A battery

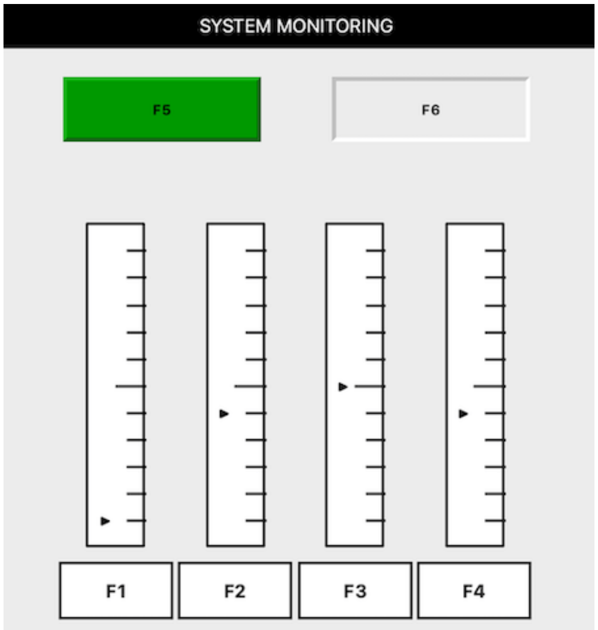
4. What information does the bar graph show? * 1 point



Mark only one oval.

- Temperature of each switch
- Temperature of each battery
- Altitude of each aircraft
- Temperature of the outside air

5. Which button in this image needs to be clicked to correct the system state? * 1 point



Mark only one oval.

- F1
- F2
- F3
- F4
- F5
- F6

6. What are the three goals of the autopilot? * 1 point

Check all that apply.

- [ ] Steering the aircraft
- [ ] Avoiding collisions
- [ ] Landing the aircraft
- [ ] Managing the batteries

7. What are the three functions of the autopilot? * 1 point

Check all that apply.

- [ ] Detect other aircraft
- [ ] Extend the landing gear
- [ ] Flip the battery switches
- [ ] Move the joystick

8. How many companies are developing air-taxi vehicles? (Choose the most correct option) * 1 point

Mark only one oval.

- More than 10
- More than 20
- More than 200
- More than 1000

9. What is the market value for these electric aircraft expected to reach by 2030? (Choose the most correct option) * 1 point

Mark only one oval.

- More than $1 billion
- More than $5 billion
- More than $10 billion
- More than $20 billion

10. What is also included in Advanced Air Mobility (AAM) besides air-taxis? * 1 point

Mark only one oval.

- Self-driving cars
- Package delivery drones
- Traditional helicopters
- None of the above

11. Where is Advanced Air Mobility (AAM) expected to revolutionize the aviation industry? (Choose the most correct option) * 1 point

Mark only one oval.

- United States
- Europe
- Worldwide
- None of the above

12. What is one of the biggest challenges for Advanced Air Mobility (AAM)? * 1 point

Mark only one oval.

- Legal obstacles to certify aircraft
- Manufacturing scalability
- High operating costs
- None of the above

Advanced Mental Model Training Assessment

<u>Note</u>: Questions 8-12 are specific for the Advanced Mental Model group.

Training

Answer the following questions based on the information that you learned from the training slides.

<u>You must score 100% on the quiz</u>, or you will have to repeat this training.

When you are done, review your score and wait for the researcher. **<u>DO NOT RESTART THE QUIZ</u>**.

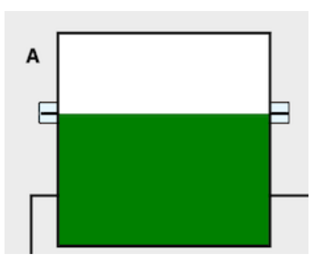
* Indicates required question

1. What are your two tasks in this study? * 1 point

Check all that apply.

- [ ] Steer the aircraft with a joystick
- [ ] Monitor the system gauges
- [ ] Manage the battery levels in each aircraft
- [ ] Avoid collisions with other aircraft

2. What is shown in this image? * 1 point



Mark only one oval.

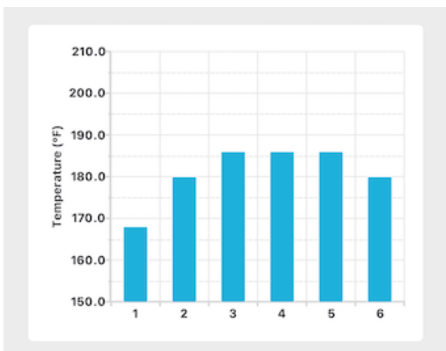
- ( ) A switch
- ( ) A battery

3. What is shown in this image? * 1 point

Mark only one oval.

- A switch
- A battery

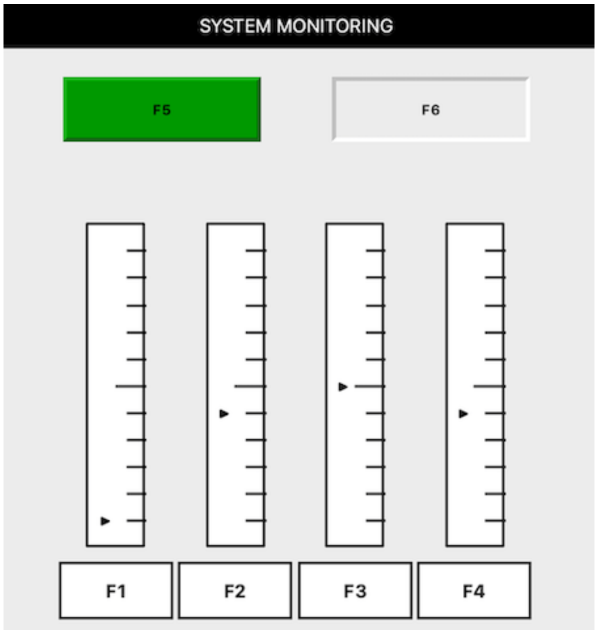
4. What information does the bar graph show? * 1 point



Mark only one oval.

- Temperature of each switch
- Temperature of each battery
- Altitude of each aircraft
- Temperature of the outside air

5. Which button in this image needs to be clicked to correct the system state? * 1 point



Mark only one oval.

- F1
- F2
- F3
- F4
- F5
- F6

6. What are the three goals of the autopilot? * 1 point

Check all that apply.

- [ ] Steering the aircraft
- [ ] Avoiding collisions
- [ ] Landing the aircraft
- [ ] Managing the batteries

7. What are the three functions of the autopilot? * 1 point

Check all that apply.

- [ ] Detect other aircraft
- [ ] Extend the landing gear
- [ ] Flip the battery switches
- [ ] Move the joystick

8. How does the autopilot determine which switches to flip? * 1 point

Mark only one oval.

- Heuristic solver
- Random solver
- Machine learning
- None of the above

9. What happens when the temperature of a switch goes above the maximum acceptable value? * 1 point

Mark only one oval.

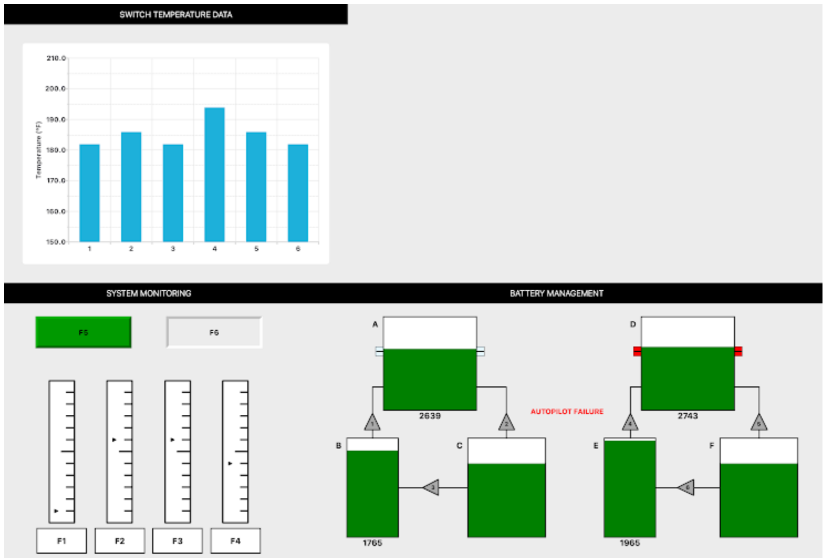
- The batteries begin to overheat
- The autopilot becomes confused about that switch
- The aircraft makes an emergency landing
- None of the above

10. Referring to the previous question, what is the maximum switch temperature that cannot be exceeded? * 1 point

Mark only one oval.

- 115°F
- 135°F
- 190°F
- 215°F

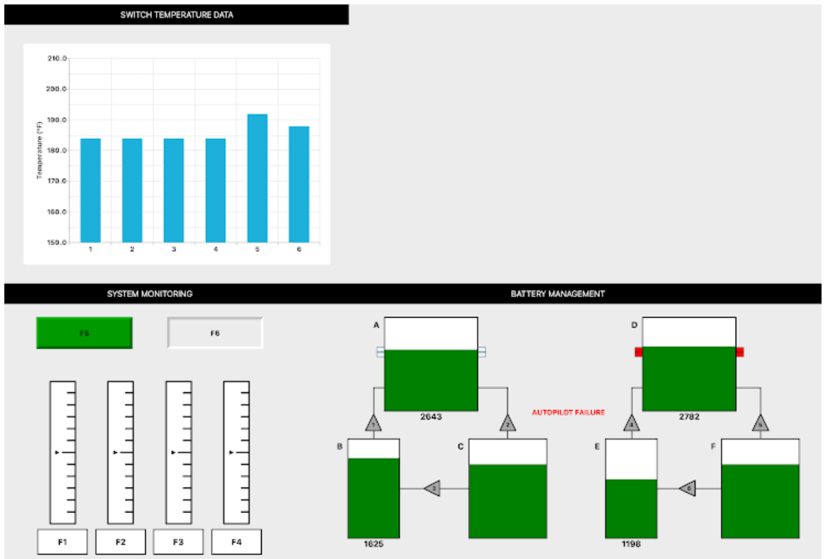
11. Which switch in this image is overheating and, as a result, needs to be reset to fix the autopilot failure? * 1 point



Mark only one oval.

- Switch 1
- Switch 2
- Switch 3
- Switch 4
- Switch 5
- Switch 6

12. Which switch in this image is overheating and, as a result, needs to be reset to fix the autopilot failure? * 1 point



Mark only one oval.

- Switch 1
- Switch 2
- Switch 3
- Switch 4
- Switch 5
- Switch 6

APPENDIX C

MENTAL MODEL ASSESSMENT QUESTIONNAIRE

Mental Model Assessment

The following questions will ask you about your understanding of the **autopilot** technology that was described in the training material. Please answer each question below.

* Indicates required question

1. The autopilot is perfectly reliable. *

   Mark only one oval.

   - True
   - False

2. Which of these is a capability of the autopilot? *

   Mark only one oval.

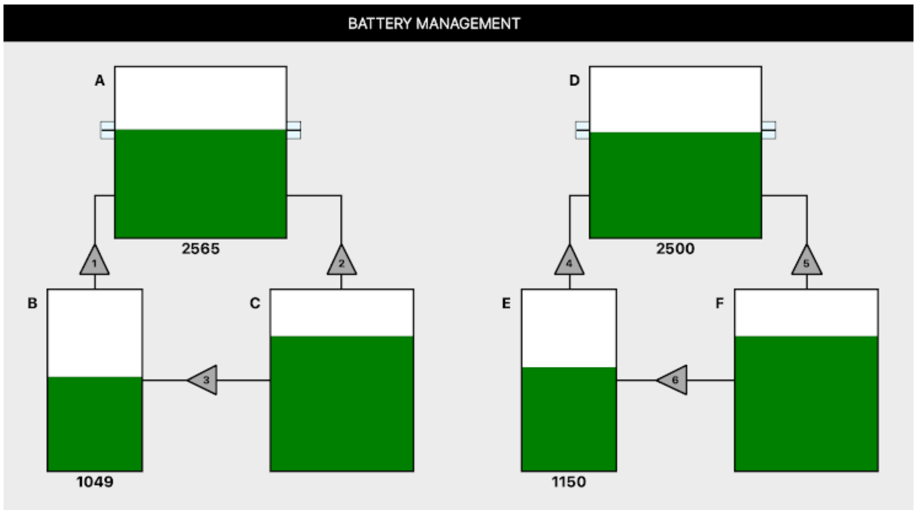
   - Landing the aircraft
   - Optimizing flight path
   - Avoiding collisions
   - None of the above

3. Which of these could cause an autopilot failure? *

   Mark only one oval.

   - High battery level
   - Switch temperature
   - Turbulence
   - None of the above

4. Based on this image, is the autopilot experiencing a failure? *



Mark only one oval.

- Yes
- No
- Unable to determine

5. Which of these is <u>not</u> a capability of the autopilot? *

Mark only one oval.

- Steering the aircraft
- Managing the batteries
- Communicating with Air Traffic Control
- None of the above

6. What does the autopilot use to make decisions? *

Mark only one oval.

- Machine learning
- Heuristic solver
- Random solver
- None of the above

7. Which of these is not a function of the autopilot? *

Mark only one oval.

- Detect other aircraft
- Flip the battery switches
- Extend the landing gear
- Move the joystick

8. Which of these could cause an autopilot failure? *

Mark only one oval.

- Severe weather conditions
- Low battery level
- Cyber attack
- None of the above

9. How is an autopilot failure resolved? *

Mark only one oval.

- Resetting the autopilot
- Replacing a battery
- Resetting a switch
- None of the above

VITA

Michael Stanley Politowicz
Department of Psychology
250 Mills Godwin Life Sciences Building
Norfolk, VA 23529

EDUCATION

2024 (Exp.)	M.S.	Psychology (Human Factors)	Old Dominion University, Norfolk, VA
2011	B.S.E.	Aerospace Engineering	University of Michigan, Ann Arbor, MI

EXPERIENCE

2019 – Present	NASA Langley Research Center	Research Aerospace Engineer
2014 – 2021	Self-Employed	Software Engineer
2011 – 2013	GE Aviation Systems	Systems Engineer (Flight Simulations)
2009	NASA Ames Research Center	Intern (Flight Deck Display Research Lab)

SELECTED PUBLICATIONS

Politowicz, M. S., Sato, T., Chancey, E. T., & Yamani, Y. (2022). Pathfinder networks for measuring operator mental model structure with a simple autopilot system. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 883–887.

Politowicz, M. S., Chancey, E. T., & Glaab, L. J. (2021). Effects of autonomous sUAS separation methods on subjective workload, situation awareness, and trust. *AIAA Scitech 2021 Forum*. AIAA Scitech 2021 Forum, Virtual Event.

Chancey, E. T., & **Politowicz, M. S.** (2020). Designing and Training for Appropriate Trust in Increasingly Autonomous Advanced Air Mobility Operations: A Mental Model Approach (Version 1). *NASA NTRS*, 43.

Sato, T., Inman, J., **Politowicz, M. S.**, Chancey, E. T., & Yamani, Y. (2023). A meta-analytic approach to investigating the relationship between human-automation trust and attention allocation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *67*(1), 959–964.

Chancey, E. T., & **Politowicz, M. S.** (2020). Public trust and acceptance for concepts of remotely operated Urban Air Mobility transportation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *64*(1), 1044–1048.

Unverricht, J. R., Chancey, E. T., **Politowicz, M. S.**, Buck, B. K., & Geuther, S. C. (2022). Eye glance behaviors of ground control station operators in a simulated urban air mobility environment. *2022 IEEE/AIAA 41st Digital Avionics Systems Conference (DASC)*, 1–6.

Politowicz, M. S., Chancey, E. T., Buck, B. K., Unverricht, J., & Petty, B. J. (2023). MPATH (Measuring Performance for Autonomy Teaming with Humans) Ground Control Station: Design approach and initial usability results. *AIAA SCITECH 2023 Forum*. AIAA SCITECH 2023 Forum, National Harbor, MD & Online.

Wing, D. J., Chancey, E. T., **Politowicz, M. S.**, & Ballin, M. G. (2020). Achieving resilient in-flight performance for Advanced Air Mobility through simplified vehicle operations. *AIAA AVIATION 2020 Forum*. AIAA AVIATION 2020 Forum, Virtual Event.