Validating the STOM Model Using MATB II and Eye-tracking

by

Garrett Zabala

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved July 2020 by the
Graduate Supervisor Committee:

Robert S. Gutzwiller, Chair
Nancy J. Cooke
Rob Gray

ARIZONA STATE UNIVERSITY

August 2020

ABSTRACT

The choices of an operator under heavy cognitive load are potentially critical to overall safety and performance. Such conditions are common when technological failures arise, and the operator is forced into multi-task situations. Task switching choice was examined in an effort to both validate previous work concerning a model of task overload management and address unresolved matters related to visual sampling. Using the Multi-Attribute Task Battery and eye tracking, the experiment studied any influence of task priority and difficulty. Continuous visual attention measurements captured attentional switches that do not manifest into behaviors but may provide insight into task switching choice. Difficulty was found to have an influence on task switching behavior; however, priority was not. Instead, priority may affect time spent on a task rather than strictly choice. Eye measures revealed some moderate connections between time spent dwelling on a task and subjective interest. The implication of this, as well as eye tracking used to validate a model of task overload management as a whole, is discussed.

TABLE OF CONTENTS

Page

LIST OF TABLES iv

LIST OF FIGURES v

CHAPTER

1 THEORIES OF MULTI-TASKING AND TASK SWITCHING CHOICE .. 1

2 THE STOM MODEL 4

STOM Attributes and Influences 8

Current Motivation 14

3 EYE TRACKING 15

4 CURRENT STUDY 19

Method 20

Results 31

5 DISCUSSION AND IMPLICATIONS 42

Eye Tracking Discussion 47

Concluding Thoughts 51

6 REFERENCES 55

APPENDIX

A PAIRED TASK QUESTIONNAIRE EXAMPLE 59

B PERFORMANCE DESCRIPTIVES 61

C SUBJECTIVE RATINGS DESCRIPTIVES 63

D SPEARMAN RANK ORDER SCATTER PLOTS 66

E EYE TRACKING TABLES 74

APPENDIX Page

F EXPLORATORY ANALYSIS MEANS 76

G MATB INSTRUCTIONAL POWERPOINT 78

LIST OF TABLES

Table Page

1. Task Pair Conflicts 25

2. Tracking Priority Ratings 33

3. Interest Ranks and Mean Dwell Times 37

4. Performance Descriptives A 62

5. Performance Descriptives B 62

6. Subjective Ratings Descriptives 64

7. Task-Focused Subjective Ratings 65

8. Priority Ranks and Mean TTE 75

9. Mean Total Dwell Time Separated by Condition 75

10. Mean Number of Dwells 77

11. Average Time Spent Per Dwell 77

LIST OF FIGURES

Figure | Page

1. Elements of STOM 6
2. MATB II Layout 22
3. MATB II AOIs 28
4. Tracking Interest Ranks & Total Dwells 39
5. Resource Management TTE & Ranks of Priority 41
6. Global Dwell Time & Highest Rank Instances 67
7. Monitoirng Dwell Time & Ranks of Interest 68
8. Tracking Dwell Time & Ranks of Interest 69
9. Communications Dwell Time & Ranks of Interest 70
10. Resource Management Dwell Time & Ranks of Interest 71
11. Monitoring TTE & Ranks of Priority 72
12. Communications TTE & Ranks of Priority 73

CHAPTER

THEORIES OF MULTI-TASKING AND TASK SWITCHING CHOICE

Multitasking is highly prevalent in several types of task environments. These environments range from the everyday driver's car sporting multimedia systems, GPS, and semi-autonomous driving features to commercial aviation flight decks (Endsley, 2017). Such technology has increased in scale which makes human monitoring and performance paramount. Human performance and choice are often affected by the specific task environment, but the effects of task choice on performance and similar relationships are still being examined (Salvucci, Taatgen, & Kushleyeva, 2006; Wickens, 2015). Whereas research has largely focused on parallel task performance and interruption management (Wickens, 2008; Trafton & Monk, 2007), multi-tasking to overcome high workload situations is of particular interest and increasingly common in everyday operations (e.g., Loukopoulos, Dismukes, & Barshi, 2009). In these conditions, the tasks are more challenging, there are often more tasks being performed than used in many experiments, and in addition, there is less environmental or rule-based structure around the multitasking and switching between them. These core differences are important because they represent deviations between laboratory-based experimentation, and real-world heterogeneous task environment examples that allow for unconstrained switching behavior.

The ways in which humans multitask can vary. The defining difference between prominent theories revolves around modes of multitasking, such as concurrent or sequential. Recently, Salvucci and Taatgen developed a model of threaded cognition which acts as a unification of theories, built onto the ACT-R cognitive architecture

(Salvucci & Taatgen, 2008) and dual-task performance model. Threaded cognition theory explores both concurrent and sequential processing related to dual tasking (Salvucci & Taatgen, 2011). The core of threaded cognition theory posits that human information processing creates goal-related threads of cognition which utilize resources serially, as a result of procedural bottlenecks (Salvucci & Taatgen, 2011).

Conceptually, multitasking and executive control are analogous to a cook preparing food in a kitchen, an analogy proposed by Salvucci and Taatgen (2011). At a cooking station, a cook has several resources available such as a stove, an oven, a chopping block, and a mixer. The cook acts as a central procedural resource whereas the tools at their disposal act as other resources such as motor or perceptual. After an order is placed, the cook must complete steps to build the desired dish. For instance, to prepare baked fish, the cook must begin by preheating the oven. Most often these steps involve a cook starting a process (e.g., preheating or boiling) and waiting for the process to conclude before they can move on with the recipe. Therefore, each step of the recipe is linked to the final dish and the states of the various resources available to the cook. If the cook has multiple orders to fill simultaneously, some processes can be completed concurrently (e.g., fish baking while pasta boils); however, resource conflicts and bottlenecks exist that will limit the amount of possible concurrent performance in other tasks as related to their concurrent demand on a particular resource. First, resource conflicts can exist when one resource is needed for two different tasks such as the oven being need for baking a fish and a cake at two different temperatures at the same time. Second, bottlenecks are also represented by the central, limited attentional resources for our cook; if two cooking processes end simultaneously for two different dishes, one dish will have to wait for the cook to attend to it after they finish the other. In other words, contention for the cook's supervisory attentional resources creates another concurrent demand that is separable from the concurrent demand on a particular resource (e.g., the oven). In relation to threaded cognition theory, the cook acts as the central procedural resource while the appliances in the kitchen represent perceptual and motor resources (Salvucci & Taatgen, 2008).

A second approach for modeling of concurrent task performance involves a theory of divided attention between concurrently performed tasks. Multiple Resource Theory (MRT) aimed to predict interference between two time-shared tasks (Wickens, 2002; 2008) by considering what mental resources were demanded by each task, and as in Threaded Cognition, consider whether there was competition or 'interference' for that resource. Task interference in MRT is thought to be the sum of three components. The first component, *resource demand*, refers to the difficulty of the two concurrently performed tasks, independent of whatever mental resources may be required to complete them. A second component, *resource conflict*, examines the range of possible overlap and sharing of demands across four dimensions of the multiple resource model; cognitive stages of processing, codes of processing, modalities, and visual channels (Wickens, 2008). The final component, *resource allocation*, provides detail about which task will suffer a performance decrement and to what extent. MRT has inherent value in predicting differences in multitasking performance across a wide range of tasks and technological designs (Wickens & McCarley, 2008; Wickens, 2008). The dimensions of the model can guide a designer to implement a certain interface feature when considering concurrent task performance. For example, is voice or tactile input best for certain tasks? Can performance of control benefit more from visual or auditory directions? (Wickens, 2008).

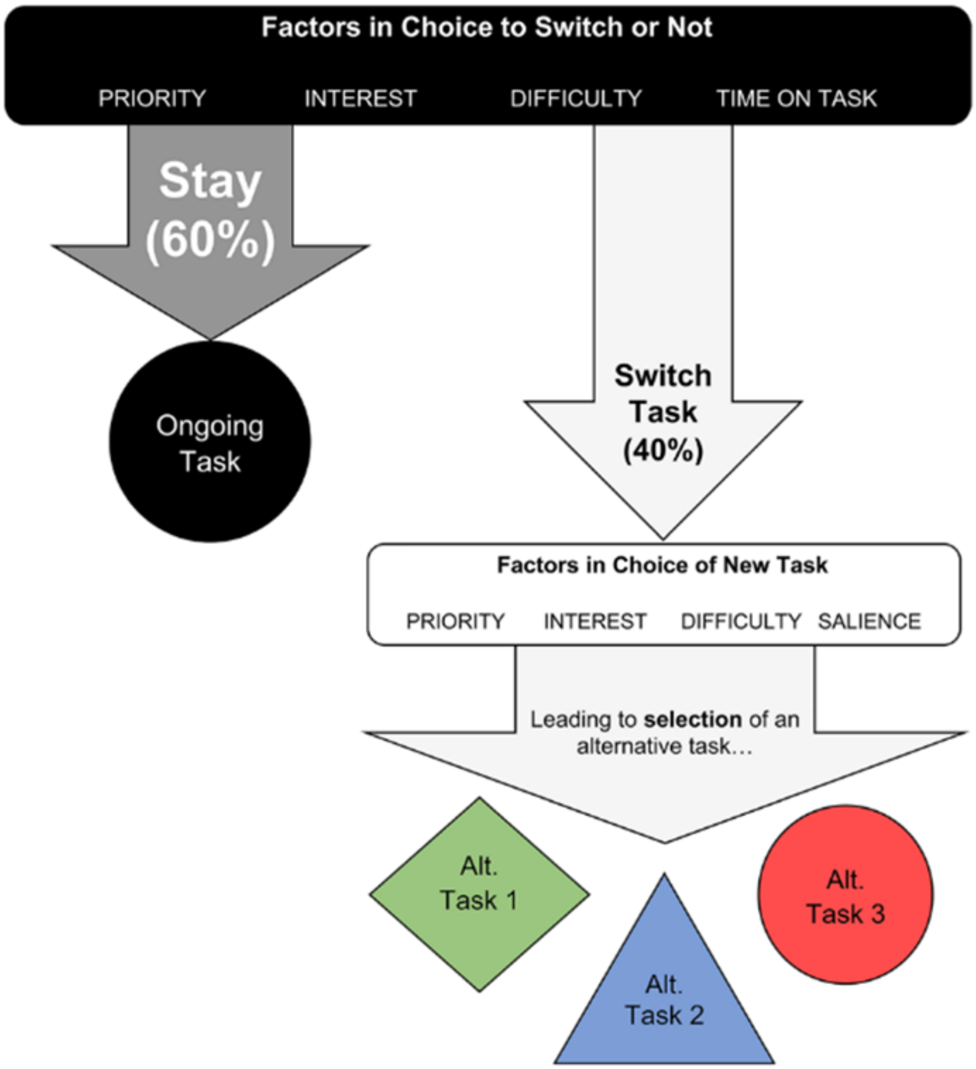
The utility of MRT and its associated model is greatest when the operator is undertaking two concurrent tasks due to increased workload (Wickens, 2008). Should workload be so low that only single-task operation is required, the model is irrelevant. But what about when workload is so great that concurrent task performance is impossible? The operator is then forced into sequential task operations and must choose to first complete one task or the other. Again, consider the cook in the kitchen analogy. Due to resource constraints, the cook can only attend to one dish if two finish at the same time. One will ultimately be delayed while the other is attended to. Choices to neglect one task, stay with a task, or switch to one of many others are intriguing due to their direct influence on real world task performance where choosing the "right" thing to do is critical (Wickens, Gutzwiller & Santamaria, 2015). Nevertheless, not all sequential model capabilities and predictions are equal. Some are limited in focus to decision making based on time, switch fluency, or time-related availability of cognitive resources (as discussed in Wickens, Gutzwiller & Santamaria, 2015). Additionally, a recent shift away from strictly resource-based theories has provided momentum for models accounting for other aspects such as perceived value, motivation, and choice (Kurzban, Duckworth, Kable, Myers, 2013).

CHAPTER 2

THE STRATEGIC TASK OVERLOAD MANAGEMENT MODEL

As mentioned above, the choice about which task to switch to among many competing tasks is potentially critical. Overloading multitask conditions are most common when technological failures occur and the operators are forced to diagnose, troubleshoot, and overcome the failure. In order to account for factors such as interest or difficulty that may influence task switching choices within sequential multitask performance, the Strategic Task Overload Management (STOM) model was developed (Wickens, Santamaria & Sebok, 2013; Wickens, Gutzwiller & Santamaria, 2015). Focusing on the decision making of an overloaded operator, the multi-attribute model aims to predict the choice of the operator in whether to stay performing an ongoing task (OT), or to switch to an alternative task (AT) during times when concurrence is impossible (See Figure 1). Further, the model attempts to predict, if a switch is made, which of the alternative tasks will be chosen; therefore, there are two choices present: the first, whether to switch or stay, and the second, which task to select having chosen to switch tasks.

Figure 1. From Gutzwiller et al. (2019), represents the fundamental elements of STOM.



In this model, the driver of choice for staying or switching, and task selection, is the concept of *task attractiveness*. Task attractiveness is based on several task-based attributes which operate uniquely in predicting the "stickiness" (resistance to switching away) of an OT as well as which AT will be selected. Attributes that contribute to "stickiness" include difficulty, priority, interest, and time on task (TOT). The attributes of an AT – which make it more or less attractive – (difficulty, priority, interest, and salience) are similar, though note that saliency is unique to influencing AT selection, while TOT is unique to influencing leaving the OT (Wickens et al., 2015; Gutzwiller, Wickens, & Clegg, 2016; Wickens & Gutzwiller, 2017). Ultimately, the combined influence of the attributes will affect the choice of whether to switch from an OT at all, and if a switch is executed, which AT will be chosen, and therefore what tasks may be temporarily neglected. In a meta-analysis (Wickens et al., 2015) of task switching literature, researchers found a 3:2 preference ratio in favor of staying on the OT, rather than switching due to a combination of the factors illustrated above. The data assisted in supplying initial weightings that could be further refined and ordered by dominance with future research and evaluation.

These five task attributes of STOM (priority, interest, difficulty, salience, and time on task) are further rooted in and reinforced by models of visual scanning and optimal information sampling that often cite the concept of attraction (Wickens et al., 2015), and were also based on a review of several literatures including task switching and interruption management. For example, fueled by visual scanning and supervisory research, a separate SEEV (Salience-Expectancy-Effort-Value) model was created to propose how visual attention is allocated to designated channels of interest (Wickens, 2015). How humans allocate visual attention is proportional to the overall attractiveness of the interface or area of interest. In the SEEV model, the key components that determine attractiveness are salience, expectancy, effort, and value. *Salience* describes the noticeable properties of a display or environment often determined by its capacity to draw attention, as well as its physical properties such as brightness, contrast, or shape. For example, a loud, auditory communication is considered to be more salient and would do a better job at grabbing attention than a colorless visual message pop-up on a display. *Effort* concerns work required to move attention from one space in the environment to another, often involving head and eye movement. Contrary to salience, effort inhibits attention movement and likelihood of fixation. *Expectancy* refers to an expected rate of change that may appear in a display, and is estimated though bandwidth (Senders, 1983). This concept of expectancy is quite relevant in automation monitoring tasks that include high vigilance demands (Wickens, 2015). The final component of SEEV is *value* and describes how tasks can hold differing levels of value or utility. Value can be related to benefit from sampling displays or can related to the relative importance of a task displayed on a screen.

Bound by the physical properties of the interface and its relation to the user, as well as what is defined as an area of interest, both salience and effort can be considered "bottom-up" factors which drive attention. Conversely, expectancy and value are guided by knowledge-driven processes and thus, can be referred to as "top-down" influences on visual attention allocation. Although expectancy is considered a knowledge driven process, it should be noted that attention may be primarily driven by more physical features such as bandwidth when the user is a novice or does not have an appropriately calibrated mental model associated with expectancy within the task (Wickens, 2015).

Though similar, the key difference between the two models (STOM and SEEV), is that SEEV predicts the attractiveness of visual areas in the environment or interface, while STOM considers the attractiveness of the tasks themselves, which may contain or include many different visual areas and information.

STOM Attributes and Influences

The current states of the attributes, relative to the model, are explored here along with some literature that feature them. The *priority* of a task can be objectively established via instructions, but also surveyed through subjective reports, and describes the general importance of a task relative to an underlying mission. Priority is analogous to the value attribute in SEEV and often directs attention towards tasks whose importance has been emphasized and may influence which one is chosen when several tasks are competing for attention (Wickens & Gutzwiller, 2017). However, priority is not as one dimensional as assigned importance and does take on different meaning depending on the model. For example, Freed's reactive prioritization model states heuristics like urgency, importance, duration, and interruption cost likely play a role in determining what tasks are attended to when the operator is temporally constrained or lacks necessary information (2000). In other words, the heuristics aim to mitigate rapidly approaching deadlines by use of time-relevant information. According to Freed (2000), the influence of priority is increased in instances of co-occurring failures, cascading alarms, or time constraints. So, there is little doubt that priority plays an innate part in task attraction but has it yet to be found as a prominent driver of task-switching behavior. In a study meant to inform the STOM model further and clarify the role of priority, researchers aimed to create scenarios in which task priority was heavily stressed through instruction (Gutzwiller & Sitzman, 2017). Although previous attempts to find an effect of priority on switch choice have been carried out in congruence with difficulty manipulations (Wickens et al., 2016 experiment 1), this one held other properties of the tasks constant (including difficulty) to further refine the role of priority. Participants completed test sessions on a task battery including four different tasks after receiving verbal instructions on what to prioritize (either one task, or all tasks equally). After the trials, participants completed paired comparison questionnaires to inform which tasks were higher in difficulty, interest, and priority. Neither task switching behavior nor time spent on a task were found to be influenced by priority instructions. Findings here have been generalized to represent the possibility of priority neglect (Wickens & Gutzwiller, 2017) and suggest people may forgo priority as an attribute guiding their attention altogether when faced with overloading, multi-task situations. However, reflecting on Freed (2000), instructions are only one way to induce priority, so these results are not conclusive, and priority remains in need of further investigation.

The *interest* attribute relates to how engaging or appealing a task may be and could be seen in a driver who becomes so occupied with their text messages that they fail to monitor the road (Wickens et al., 2016). Here, interest in texting overrides the priority of safe driving. Work cited (Spink et al. 2006) in an earlier meta-analysis posits interest as an attribute that can influence engagement in an applied, task-switching domain (Wickens, Santamaria, & Sebok, 2013). Specifically, Spink et al. were interested in human-computer interaction and how people construct and manage their information when tasked with exploring set topics (2006). Participants were told to adopt any search technique they felt comfortable with, and to describe their process of information searching for analysis. Participants also rated the attributes pertaining to their technique in terms of difficulty, interest, and prior knowledge. Results indicated that several factors influenced how participants searched for information, and in particular their level of personal interest was most predominant and considered to be a major factor in 45% of the subjects. Difficulty, or ease of finding information in this case, was also important, being cited by 20% of participants. Together, the influence of interest was clear in the test of an applied, task-switching domain; however, it should be noted that it was not the only factor.

Salience, akin to the salience described in SEEV, characterizes how attention-grabbing the arrival of a task is. For example, does the task arrive based on a visual alarm firing, an auditory alert or, is it a task that is only performed when an operator remembers to do it? These all have different salience because of the sensory attributes involved. In other words, a loud auditory alert has higher salience than a visual status notification which has higher salience than an operator's prospective memory to perform a task in the future (Dismukes, 2010).

Difficulty is somewhat analogous to effort, in that the level of mental workload and demand of a task can critically influence switch behaviors. Previous research, including the meta-analysis above, revealed a bias of choosing easier tasks over more difficult ones (Wickens et al., 2016 Experiment 1). In other words, difficult tasks hold a more negative weight as AT and so, they are generally switched to less. However, if a difficult task is switched to, there is evidence to suggest the weighting changes (Wickens et al., 2016, experiment 2). For example, researchers manipulated difficulty in a STOM validation study involving astronaut simulation environments to further study the influence of difficulty on task switching. Participants were responsible for two tasks: a process-control task which largely involved management of systems with the help of automated aids, and a robotic-control task which involved repair tasks, and could be either automated or manual depending on the difficulty manipulation. Workload varied by the inclusion of decision support and automation of the robotic-control task. The measures of attention and task switching between the sequentially performed tasks were acquired by both a head tracker and control related activities that forced the participant to focus their attention to one of three displays encompassing a 120º viewing angle. The results indicated that, again, priority does not influence attention allocation. More importantly, participants performed more difficult tasks for longer than easier tasks which suggests overall switch resistance while completing a difficult task. This means that difficulty may play an important role in determining whether a person stays with a given task, with the probability increasing when that task is more difficult.

The fifth and final attribute – *time on task (TOT)* – is shrouded in considerable complexity. Evidence suggests that TOT has been tethered to accumulating fatigue, cognitive effects of unpredictable perturbances (Sheridan, 2007) as well as resource depletion (Kurzban et al., 2013; Baumeister, Vohs & Tice, 2007) causing an increased likelihood of task switching. Put simply, the longer an operator stays performing a task, the more likely a break will be needed due to resource depletion and as a result, they are increasingly prone to a task switch. Again, a cognitively-based theory proposed by Sheridan suggests that a dynamic alternative task (AT) that is suddenly influenced by uncertain circumstances would likely prompt a task switch (decreased switch resistance) because more time spent on the OT could make matters worse surrounding the AT. In other words, increased TOT on the OT while a perturbance occurs in an AT will likely raise questions as to the state of the AT and perhaps the entire system. This would suggest an increased likelihood of switching with more time spent on the OT.

Contrary to the monotonic TOT effects mentioned above, the following effects occur in a more periodic fashion and depend on the operator to be in a specific phase of a task’s performance to be useful in prediction. First, complex tasks may favor higher TOT due to high working memory demands (Gutzwiller, Wickens, & Clegg, 2016), which may act as a form of switch resistance (Wickens & Gutzwiller, 2017). Here, switching away from a high-memory task could exacerbate working memory demands, influence forgetting of critical contents (such as a verbal message), and force restarting the task upon return; thus, avoiding switches during high-load tasks may periodically result in greater TOT in such situations. That is not to be confused with task inertia which implies tasks will be prioritized because the operator is nearing an end point and demand has accumulated over the task (Gutzwiller, 2014). This inertia effect ensures that operators are using their mental resources effectively, as leaving then returning to a task may create a cost to resume it (and refreshing memory may not be possible, as the original information has passed). Tasks in which inertia is most common are high in working memory maintenance of an OT. Ultimately, the utility in determining what features of an environment or task may lead to operators staying or switching is useful for design because of its connection with task performance and its use in preventing extreme aversion to staying or switching. Although the underlying role of TOT with relation to task switching is not the entire focus of the current work, it is connected to visual sampling which has not yet been explored in the context of the STOM model. Additionally, previous validations of the STOM model have mainly used performance of tasks as a proxy for where operators are allocating their attention (Gutzwiller, 2014; Gutzwiller, Wickens, & Clegg, 2016, Gutzwiller & Sitzman, 2017). Operators seldom, if ever, choose to perform a task without actively allocating their visual attention to that task and its space in the environment. So, a remaining question is: How do operators allocate their attention between discrete task performances or OT performances and, ultimately, how much bearing does that have on task choice and switching?

Current Motivation

Time spent visually sampling remains an unexplored matter related the current experiment's predictions related to the STOM model. Previously, utilizing a multi-task testbed, a continuous performance measurement was easily taken of a compensatory tracking task, because it necessitated constant input from a participant (Wickens & Gutzwiller, 2017). However, other tasks such as the monitoring task and the fuel management task, which may still require continuous vigilance by the participant, only provide discrete performance input. In other words, tasks such as system monitoring and fuel management may benefit from time spent visually sampling, even when the task itself not being directly performed. This sampling has not been previously measured or understood in the context of the model. By proposing the measurement (in the form of eye tracking) of continuous visual attention allocation, we can capture visual attentional switches that do not manifest into behaviors, but may provide additional insight or predictive value into task switching and task management. Considering the above, it is the equivalent to merging the SEEV approach to visual allocation prediction, with the STOM approach to task selection and prediction.

CHAPTER 3

EYE TRACKING

Extending the validation of STOM with the addition of eye tracking requires identifying particular eye tracking measures to be used during a replication experiment. More specifically, *position* measures and *latency* measures are most applicable to the current research for two reasons. First, to capture visual attentional switches, eye positioning must be recorded (Holmqvist, Nyström, Andersson, Dewhurst, Jarodzka, & van de Weijer, 2011). Second, position and latency measures best support the objective of finding connections between subjective ratings of task attributes in the STOM model, and actual behavior. For instance, if a participant rates a task as more difficult than another, eye tracking data should assist in testing that claim through the use of the *dwell time* measure. Thus, people should be shown to spend more time looking at a more difficult task if they are spending more attention on it. Similarly, if a participant rates a task as higher priority than another, and thus garners faster responses to events as they arise, eye tracking data should assist in testing that claim through the use of *latency* measures.

Position duration measures are key in determining how long a participant has gazed in a certain position (Holmqvist et al., 2011). This position is always either that of a fixation location, or more broadly within an "area of interest" (AOI). Multiple dwell times can be measured, but for the current purpose, fixation duration is not as useful as dwell time duration. This is because fixations are limited to a single position, whereas dwell time is the duration of time spent looking anywhere within an AOI. AOIs can be defined easily for any eye tracking study using a screen or display; for the current research, AOIs are areas of the MATB II display that correspond uniquely to each specific task (Figure 3). In some instances, AOIs can be intricate and have several features for participants to gaze upon. As dwell time is often comprised of several fixations within an AOI, it is logical to have collected dwell time duration, rather than tracking specific fixations.

AOIs are bounded as the visual area containing each unique task within the MATB simulator (e.g., monitoring AOI, resource management AOI). The specific position duration measure that was used is total dwell time duration which, along with the total number of dwells, can reveal average time spent per dwell. The former acts more like an aggregate measurement and sums the amount of dwell time spent in a certain AOI over a whole trial. Research has revealed that difficultly obtaining information from an interface can produce longer dwell times (Goldberg & Kotval, 1999). Though the cause of difficulty may stem from poor interface design, there is at least a feint connection between difficulty extracting information from a display, general task difficulty, and associated increased visual dwell times. Finally, when tasked with choosing the most appealing object within unfamiliar stimuli, researchers have found that dwell times gradually rose on the most attractive object up until the object was ultimately chosen (Shimojo, Simion, Shimojo, & Sheier, 2003). Although not directly related, the research is conceptually linked to an attribute such as interest and therefore, is useful to consider when examining task switching with visual sampling. Such position duration measures paired with existing subjective measures for assessing operators task attribute ratings and preferences could assist in populating STOM attribute weightings by observing task switching behavior in the context of TOT.

Latency measures were also used in the present study and offered similar benefits as position duration measures. Latency refers to a direct measure of time delay (Holmqvist et al., 2011), as a measurement of the elapsed time between the on- or offset of an event, and the on- or offset of another. Several latency measures within eye tracking exist (e.g., saccadic latency, smooth pursuit latency, latency of the reflex blink, pupil dilation latency); however, time to entry (TTE; also known as entry time) to an AOI is most useful for the purposes of this work. TTE of an AOI measures how soon after a stimulus onset a subject's eyes enter an AOI. In general, shorter entry times to an AOI suggest greater efficiency in detecting a stimulus in question (Holmqvist et al., 2011). With this in mind, TTE can be particularly beneficial in examining how participants approach overloaded multitask events. In this study, these are referred to as event conflicts or events that occur nearly simultaneously that participants must respond to. In particular, participants in prior work on STOM could have noticed events, but simply chose not to perform them (or perhaps forgot to perform them). The addition of eye tracking measures could reveal this type of behavior in our paradigm.

Various factors of the tasks themselves are assumed to influence the measures we are interested in above, particularly TTE. For example, Eberhard, Spivey-Knowlton, Sedivy, and Tanenhaus (1995) found hearing speech can lead to earlier entries into an AOI, when the spoken speech corresponds well and creates a semantic connection. Similarly, the general conspicuity of the elements of an AOI can lead to shorter entry times; in a study examining visual attention and recall of conspicuously designed warning labels, researchers found entry times to be significantly faster for more visible designs (Krugman, Foxer, Fletcher, Fisher, & Rojas, 1994). A similar connection may exist between tasks determined to be more salient, and attention allotted to them in our paradigm, particularly during overloading multitask situations (in the case that salience is observable and definable, at least). For the STOM model, the attribute of task salience differs slightly from a primarily visual definition (a signal-to-noise measure of contrast between the target and surrounding environment relative to stimulus dimensions like luminance, motion, color, depth, and orientation; Nothdurft, 2000). Salience in the STOM model has been defined in prior work as a generalized hierarchy established based on prior studies and assumed as such; auditory events are more salient than visual (Lu et al., 2013), which are both more salient than prospective memory to attend to some event in the future. So, auditory events are believed to have shorter (faster) time to entry than other, less salient events.

CHAPTER 4

CURRENT STUDY

The current study had two main purposes. First, it sought to replicate prior results from Wickens et al. (2016, Experiment 1) to validate the STOM model and better populate the attribute weightings. Second, the study aimed to address a limitation of existing STOM validation studies by providing the ability to capture visual attention switches made by participants; previous efforts had relied on performance inputs to determine task switches only. In other words, though STOM has received some validation, it has largely been via tracking what participants do, which may co-occur or occur in tangent with where and when they look and direct visual attention within the interfaces. With the addition of eye tracking measures, the current study will add to the existing literature from a novel perspective using dwell time duration and TTE to AOIs.

The study tested six main hypotheses related to both the experimental conditions, subjective rating questionnaires, and associated eye tracking measures. In accordance with Gutzwiller et al. (2014), the first three hypotheses listed mirror those of previous research:

1) Fewer switches should occur in difficult tracking conditions compared to easy tracking conditions because task switching is believed to be cognitively effortful.
2) A more difficult task should invoke proportionally fewer switches to it, whether difficulty is measured or manipulated.

3) A task with a higher priority should reveal fewer switches away from it when it is the ongoing task (evident by a stay preference) and may lead to more switches toward it if it is an alternative task (a sign of attractiveness).
4) A task with higher subjectively rated interest should garner a greater amount of visual dwell time.
5) Tasks with high rated priority should incur lower entry time to its defined AOI.[1]
6) The difference in the tracking task dwell time between the high and equal priority groups will be greater when the tracking task is more difficult.

Method

Participants

Based on a prior study and power analysis (Gutzwiller et al., 2014) a sample size of 80 was determined to be suitable. However, due to the global spread of SARS-CoV-2 (sever acute respiratory syndrome coronavirus 2), participation was severely hampered. Twenty-seven students participated either by volunteering for partial course credit or were recruited as paid participants and received $12 per hour (of the 27, three were paid participants). Participants who were paid were not incentivized to perform better than others. Most participants were recruited through a subject pool associated with Arizona State University in addition to alternative avenues such as personal networks, and

[1] Hypotheses 5 and 6 were proposed as dependent on the paired conflict event data. However, due to conflict events not occurring exactly at the same time in MATB, proper analysis could not be conducted. Instead, both hypotheses were assessed using more general event data. Further explanation can be found in the *Pair Conflicts* subsection.

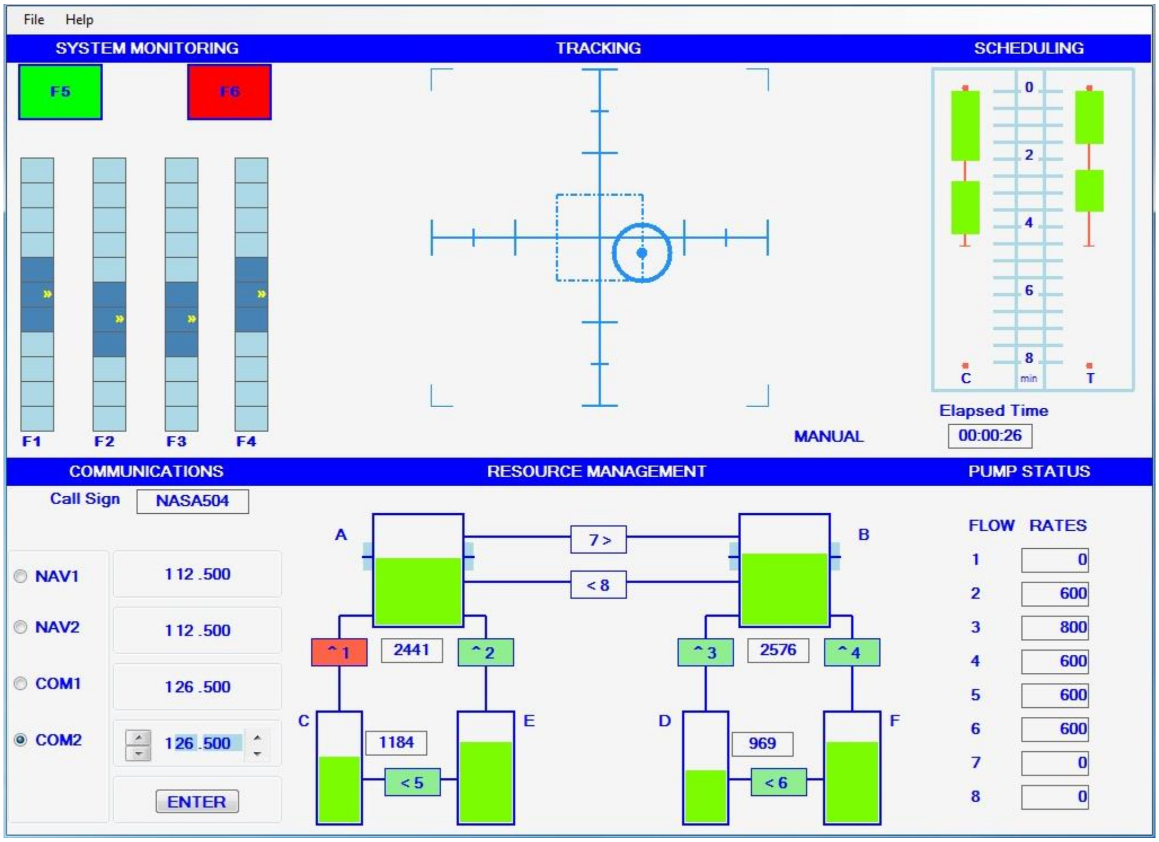
physical flyers with the encouragement of paid participation. Recruited participants were then compensated with course credit in accordance with subject pool policies.

Materials and Design

To further validate STOM, contribute to attribute weightings, and replicate the methods of previous research (Gutzwiller et al., 2014) the MATB II suite (Santiago-Espada, Myer, Latorella, & Comstock, 2011) was utilized. Originally designed to mimic cognitive task conditions pilots experience during flight, the MATB is a customizable, computerized platform that effectively simulates multi-task demands (Santiago et al., 2011). The tasks do not require aeronautical knowledge but still fulfill the more important requirement of generating high multi-task demands in a measurable environment. Figure 2 illustrates the main interface of MATB II including its four main subtasks.

The four main subtasks included the system monitoring task, tracking task, resource management task, and the communications task. Below, each subtask is explained in the context of the current study.

Figure 2. A screenshot of the Multi-Attribute Task Battery-II (Santiago-Espada et al., 2011). Beginning in the upper left and moving clockwise, the system monitoring, tracking, resource management, and communication tasks are shown.



System Monitoring. The system monitoring task was a visual supervisory task that included three different subtasks. Participants monitored four scales and the two lights above them. The indicators in the scales (dark blue sectors with yellow chevrons) fluctuated normally around the center but could indicate a problem requiring a response by shifting for a short period of time to either the top or the bottom area of the scale. Responses were made with a mouse, clicking the corresponding scale demanding a response. Once a scale was clicked, the indicator returns to normal fluctuating behavior around the center, until another event is triggered. The light on the top left (green) is on if the light appears green. Should the light turn off (leaving the light colorless), a click response with the mouse was required and would turn the green light back on. Similarly, the light directly to the right (red) was normally off (colorless). If the light turned red, a response with the mouse was required and would result in turning the red light ‘off’ to its normal operating status. All in all, a normally functioning system state included a green light, a colorless right light, and center-fluctuating scale indicators. Participants responded to both the lights and indicators and reaction time and accuracy in response were recorded as outcome measures. Note that each of these subtasks has a ‘timeout,’ meaning that responses to the event are available for ten seconds before the system returns to a normal state. (Thus, if a participant does nothing for the ten second period of time, the response is recorded as incorrect, and the maximum RT – 10s – is assigned to the event).

Tracking. The tracking task was a continuous one, requiring the participant to use a joystick and control a circular reticle. The reticle constantly moved with various amounts of displacement according to a sine wave function driven by the MATB II simulation. The goal was to keep the circular reticle centered on the crosshairs in the middle of the dotted square (see Figure 2). A greater amount of reticle displacement required a greater amount of manual input via the joystick, which is further influenced by settings in the MATB platform (i.e., greater bandwidth). Performance error was measured by the distance between the current location of the reticle and the centered location, recorded at a rate of 1 HZ. The tracking task was important in the experimental manipulations in two ways. First, difficulty was manipulated within participants across two trials using the tracking task; and second, the attribute of task priority was also manipulated through verbal instructions emphasizing (or not) the priority of the tracking task as a between participants manipulation.

Communications Task. The communications task was located in the bottom left corner of the interface and represented a discrete task type with several steps. The task itself involves the presentation of auditory, simulated air traffic control radio transmissions, which contain instructions on radio and frequency manipulations that can be made in the interface (Figure 2). Radio transmissions consist of both own-ship (relevant) and other-ship (irrelevant) radio calls. The participant is only required to respond to the own-ship calls, and should ignore all others. The transmissions consist of a callsign (NASA 504) and instructions to switch to one of four possible radios, and then to alter the frequency of said radio. Together, a radio call might sound like this: "NASA five zero four, NASA five zero four, tune your nav one radio to one one four point five seven five." Audio requests for the task were presented through built-in speakers of the laptop at a volume level of 60%. Participants used the mouse to select the radio and set the proper frequencies before hitting a final "enter" button. Performance on this task was measured by reaction times and completion accuracy for both radio and frequency.

Resource Management Task. The fourth and final task was the resource management task. The objective of this task was to maintain the correct and constant amount of fuel in each of two critical fuel tanks (Tanks A & B; see Figure 2.). To successfully sustain performance in this task, the participant must have managed fuel flow from several auxiliary tanks (lettered containers) to the critical tanks using fuel pumps (numbered paths between tanks). These pumps have differing rates of fuel flow and can fail (indicated by the fuel pump switch turning red; e.g. pump 1). Should a fuel pump fail, the participant must reroute fuel to maintain the desired level in the two critical fuel tanks. Again, the overlying goal was to maintain the fuel level around the blue indicators on either side of both the critical tanks. Performance was measured through accuracy in terms of overall error derived from the difference in each task from the target level, at a rate of 0.1 HZ

Pair Conflicts. To measure task switching preference from the continuous OT (Tracking) to two potential ATs, task event pair conflicts were injected into both test trials. In summary, a *pair conflict* involved a stimuli or status change event in two different tasks of the MATB simulation in a nearly simultaneous fashion (Gutzwiller et al., 2014). There are three different types of pair conflicts and their induction time as well as their incidence order varies randomly across trials. The types of pair conflicts were created so that each task was paired with the other two remaining tasks (Table 1)**.**

Table 1. Representation of task pair conflicts in each test trial (from Gutzwiller et al., 2014).

	TRK	MON	RMAN	COMM
TRK	x	x	x	x
MON		x	2 conflicts	2 conflicts
RMAN			x	2 conflicts
COMM				x

Each type of pair conflict occurred twice during each test trial, for a total of 6 conflicts per trial. Depending on task attributes and relative attractiveness, one task of the pair should be chosen first to be completed, while the other is either neglected or forced into a possible queue to be performed following the first. It is anecdotally very difficult to attempt to perform both at the same time (concurrent performance) and with the additional ongoing tracking task results in a potential overload situation.

Initially, hypothesis five predicted higher priority tasks would be led to lower time to entry to its AOI after a conflict event. However, due to restrictions within the MATB II software, truly simultaneously occurring events are not possible. To assess TTE during conflicting events, only the first event that fired was considered for analysis. Hypothesis six also previously relied on simultaneously occurring conflict events and predicted more dwell time would be associated with the more difficult of the two conflicting events. Using the same justification regarding H5, the sixth hypothesis abandoned conflict events entirely and currently examines interactions between priority, difficulty, and dwell time.

Priority Induction Conditions. Acting as one of two main influential parameters of the current study, priority was manipulated to further explore its influence and to replicate the methods of Wickens et al. (2016, experiment 1). To induce two different priority conditions, a between subjects design was carried out through verbal instruction. In one condition (equal priority), participants were told to equally prioritize all MATB tasks while in the second condition (tracking priority), participants were instructed to prioritize the tracking task while performing all tasks as best they can. Participants were only to execute the tasks with a single, dominant hand to avoid parallel responses which could lead to concurrent multitasking.

Difficulty Induction Conditions. Acting as the second variable of interest, difficulty was manipulated to further explore its effect with regard to STOM and previous research. Task difficulty was manipulated using the Tracking task within participants across two trials (one easy, one difficult). Tracking difficulty was manipulated through the MATB software to increase the update rate of the tracking task reticle (additional movement bandwidth) and the responsiveness of the joystick was lowered in terms of response speed to increase the difficulty. The difficulty manipulation also mimicked those made in prior work (Gutzwiller et al., 2014; 2015; 2016; 2017; 2019).

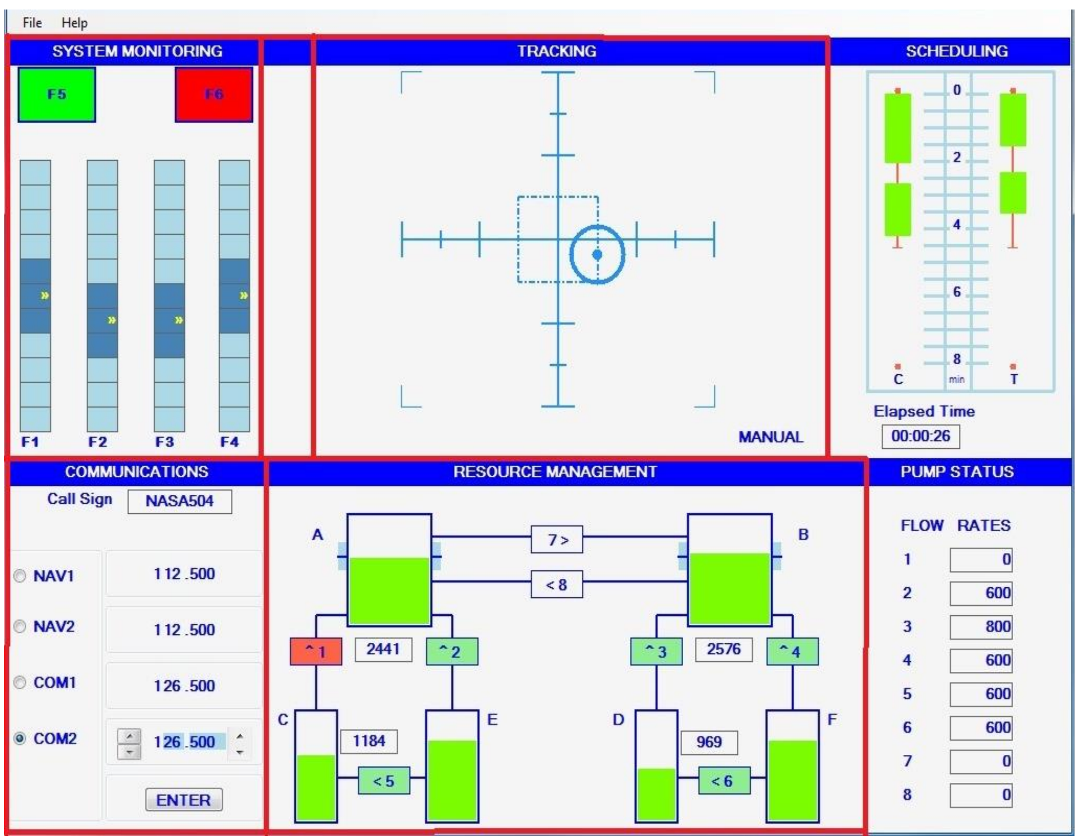
Eye Tracking. Continuous recording of eye movement was employed for the current study to address attentional actions that may not manifest as measurable behaviors through the MATB interface. Eye movement was recorded through the Tobii X2-60 Eye Tracker with a sampling rate of 60HZ. Associated eye movement data was collected via iMotions software for data analysis. The particular eye tracker is sufficient for the current study due to the necessary measures not requiring high sampling frequency or advanced data outputs (e.g., pupillometry). Specific iMotions measures used were gaze dwell time and number of fixations. Entry time to AOIs (also known as time to first fixation) were calculated manually using the iMotions replay feature which allowed for manual analysis of eye movements in terms of thousandths of a second. All eye gaze data was time-synced to the MATB experimental trials and assessed according the defined AOIs.

Areas of Interest (AOIs). In short, AOIs delineate regions in the stimulus and encapsulate what is "interesting" to further characterize various analyses of eye-movement data (Holmqvist et al., 2011). As opposed to data samples that strictly involve eye movements such as saccades, fixations, or smooth pursuits, AOIs add an additional layer of density by including stimulus content that can closely relate to hypotheses posed by the researchers. In the current study, the overall stimulus that was comprised of AOIs was the MATB interface. In turn, data of simple eye events transformed into data of AOI events. Further, the AOIs in the current study were defined by the four tasks of the

MATB interface and were specifically demarcated through the iMotions software. Below, (Figure 3) is an approximation representing the AOIs and their borders.

Generally, AOIs should not overlap to avoid duplicate measures, and tricky-to-interpret data (Holmqvist et al., 2011). For the current study, there are no overlapping AOIs but, there are AOIs that almost share a border. For the purpose of illustration, the lines demarcating the AOIs in Figure 3 seem to share borders but, the iMotions software allowed for much finer AOI creation. Put simply, no AOIs shared a boarder. In the case of the system monitoring task and the tracking task, these two AOIs share ample white space in the interface that is not of interest and thus, was not included in any AOI.

Figure 3. The AOI borders are outlined in red. The borders will not actually be visible to the participant and, rather, are predefined for the use of eye tracking measures.



Subjective Measures. In addition to eye tracking, subjective measures of attribute weights were gathered to determine which tasks were more difficult to perform, of higher priority, and more interesting (Gutzwiller et al., 2014). The survey (see Appendix A) is a paired comparison rating scale that equally contrasts all tasks for attribute weight assessments. For example, monitoring and resource management are compared and participants are asked to rate on a continuum which task is (a) more difficult, (b) more interesting, and (c) higher priority.

Differences from the Replicated Study

The current study differs from Wickens et al., (2016, experiment 1) in that the task comparison survey ratings were completed at three points throughout the study: (1^{st}) after the 2-minute training trial, (2^{nd}) after the first test trial, and (3^{rd}) after the second and final test trial. The original study only used this rating scale after the test trials were completed. The additional measure used in the current study was meant to establish a baseline of subjective task weights and allow for measurement over the course of the experiment. The initial measure taken after the practice trial but before the priority induction, allowed researchers to see if certain groups of the participants began as equivalent or not.

The second major difference is the inclusion of the eye tracker which adds some experimental complexity but, was not expected to influence results outside of adding some time for calibration and recalibration between the two test trials.

Headphones were not used in this experiment as in prior work, in favor of speakers, but again, this difference was not expected to influence results. In particular, headphones were used in prior experiments to provide participants with more isolation given they were often run in parallel with other participants, whereas in the current experiment the participant was by themselves with only the experimenter present and no competing noises.

Procedure

Following the consent procedure, participants were assigned a participant number according to their order of participation in relationship to previous participants. At the same time, the participant was randomly assigned to one of two conditions (Equal Priority or Tracking Priority) and a counterbalanced order of two test trials which varied in tracking difficulty (difficult-easy, or easy-difficult).

Following assignment, the participant sat at the experimental station and read through a self-paced training presentation (see Appendix G). The presentation explained the MATB II tasks in detail and how to perform each one. Next, a brief 2-minute training trial in which the participant performed the four tasks in the MATB suite all operating at the same time occurred. Following the training trial, the participant was asked to complete the first brief, subjective paired-comparison rating scale in which task pairs were compared on difficulty, priority, and interest.

The participant was then familiarized and calibrated on the eye tracking equipment. The calibration process included properly positioning the participant in space according to the eye tracker system feedback. Participants completed the calibration by following a moving white dot across the screen with only their eyes. Calibration quality of “Excellent” according to the system was the goal but calibration quality of “Good” was also acceptable. No bite bar or chin rest was used in this study, and this exclusion was not expected to majorly influence the results. The iMotions software used to analyze the eye data visually displayed proper head and eye placement relative to the eye tracker prior to calibration. Participants were briefed on this and were shown the optimal distance from the eye tracker and head positioning relative to the eye tracker.

Next, depending on their group assignment, the participant was introduced to the test portion of the experiment and was either told to perform all tasks as best they can (Equal Priority) or to prioritize the tracking task (Tracking Priority). Then, the participant completed one 10-minute test trial in the MATB platform with the tracking task being continually active, intermittent single task conflicts being activated, and 6 paired event conflicts occurring (either easy or difficult tracking condition, based on counterbalance order). Following the conclusion of the first trial, the subject completed the second task comparison rating scale questionnaire. It is important to note that participants were given short breaks between test trials in addition to being recalibrated on the eye tracker to ensure accurate collection.

Following recalibration, the second and final test trial was completed. Again, the test trial comprised of a 10-minute simulation containing a continuous tracking task, intermittent single task conflicts being activated, and 6 paired event conflicts. The two test trials varied in tracking task difficulty (easy or difficult) and were counterbalanced. After the second test trial, the participant was asked to complete the final task comparison rating scale. This concluded the experiment.

Results

Outliers greater than 3 SD from the mean were moved from all analyses. All analyses were conducted on the 27 participants unless noted. All assessments of Shapiro-

Wilk test of normality revealed normality unless otherwise specified. Mauchly's test of sphericity, where warranted, was conducted and none were significant (all $p > .05$). For hypotheses involving three-way ANOVAs initially including counterbalance order, no significant effect was found for order. All ANOVAs results following have order removed as a factor.

Task Performance

An operator's ability to sequentially switch tasks in high workload situations does have bearing on performance (Wickens, 2017). However, the STOM model's focus is the decision to switch in addition to the TOT spent rather than classification of performance. Still, broad performance measures were taken in the current study to assess certain manipulations such as difficulty. Also, to ensure participants were not abandoning tasks and therefore altering the task environment to include less tasks to attend to, performance was gauged generally. It appeared that all tasks were performed across both trials and both conditions (see task accuracies and reaction times in Appendix B). Because of the central nature of tracking task performance to the hypotheses (being manipulated for both priority, and difficulty) a simple analysis was run on this performance data. One outlier more than 3 SD from the mean was removed from the comparison. Tracking performance was measured in terms of root mean squared deviation. To assess any differences in tracking performance, a 2 (tracking difficulty) x 2 (priority assignment) repeated measures ANOVA was conducted. The tracking difficulty manipulation was successful in that participants fared better (less error overall) in the task when it was easy (M = 19.87, SD = 1.03) than when it was difficult (M = 39.51, SD = 1.83; F (1, 24) = 151.49, p = .000, η_p^2 = .863). No between-subjects effect of priority assignment was found, F (1,

24) = 71.35, p = .359, η_p^2 = .035 and there was no interaction between difficulty and the priority assignment F (1, 24) = 1.32, p = .262, η_p^2 = .052. For more descriptive statistics regarding performance, reference Appendix B.

Task Attributes Questionnaire

Paired task comparison ratings were taken for three task attributes relevant to STOM (priority, difficulty, and interest) at three times during the experiment. Of initial interest here were ratings given on priority after the priority manipulation (surveys 2 and 3), as a manipulation check on the instructions actually establishing subjective prioritization. The tracking priority group did rate the tracking task as higher priority (*M* = 5.5, *SD* = 2.66) compared to the equal priority group (*M* = 2.3, *SD* = 3.71) when averaged across the second and third administrations ($t(52) = -3.48$, p = .001). When each administration is compared – separated by condition – the mean priority ratings for the tracking task can be found in Table 2.

Table 2. Subjective questionnaire priority ratings for the tracking task separated by priority condition.

	Equal Priority	Tracking Priority
1st Administration	4.9	1
2nd Administration	2	5.7
3rd Administration	2.6	5.2

To assess for any differences between priority groups before the priority manipulation was carried out, independent sample t-tests were conducted with each tasks' means from the first questionnaire administration. For the tracking task, the equal priority group (*M* = 4.93, *SD* = 3.99) was significantly different from the tracking priority group (*M* = 1, *SD* = 3.65; $t(25) = 2.66$, p = .013), suggesting the two groups did not start as equivalent.

Additionally, for the communications task, the equal priority group (*M* = -3.36, *SD* = 5.94) began as significantly different from the tracking priority group (*M* = 2.54, *SD* = 3.95; *t*(25) = -3.01, *p* = .003). Examination of the two groups regarding the remaining tasks reveal no significant differences. All other descriptives regarding the subjective questionnaire data can be found in Appendix C.

Task Switching

Task switching was measured by assessing participants responding to any of the four tasks of the MATB II simulation. Hypothesis 1 predicted fewer switches in difficult conditions compared to easy. A 2 (tracking difficulty) x 2 (priority assignment) repeated measures ANOVA was run to determine the effect of tracking difficulty on amount of task switches performed. There were fewer switches in easy (*M* = 48.19, *SD* = 10.75) as compared to difficult (*M* = 52.93, *SD* = 10.84) and no significant effect was found for difficulty, $F(1, 25) = 4.15$, $p = .052$, $\eta_p^2 = .14$. The between-subjects effect of priority assignment was not significant, $F(1, 25) = .025$, $p = .877$, $\eta_p^2 = .001$. Finally, there was no significant interaction between difficulty and priority assignment, $F(1, 25) = 2.13$, $p = .157$, $\eta_p^2 = .079$. The results provide evidence against H1, with fewer switches occurring in easy tracking trials compared to difficult trials, although not in a significant way.

To assess H2, a one-way, repeated measures ANOVA was conducted to determine the effect of tracking difficulty on the amount of switches *to* the tracking task as opposed to from it. A main effect of tracking difficulty was found but not in the direction that was hypothesized. Participants switched to the tracking task *more* when it was difficult (*M* = 24.41, *SD* = 5.83) than easy (*M* = 19.85, *SD* = 6.32; $F(1, 26) = 13.28$,

$p < .001$, $\eta_p^2 = .338$). Refuting the second hypothesis, participants switched to the tracking task more when it was difficult than easy.

Hypothesis 3 predicted a higher priority task to have fewer switches away from it (as an OT) and more switches to it (if an AT). To evaluate H3, an additional a 2 (tracking difficulty) x 2 (priority assignment) repeated measures ANOVA was run to determine the effect of *priority* assignment on switches *away* from the tracking task. No effect of priority was found, $F(1, 25) =.156$, $p = .696$, $\eta_p^2 =.006$. There was, however, a significant effect of difficulty with more switches away from tracking occurring in difficult trials ($M = 24.70$, $SD = 5.92$) than in easy trials ($M = 19.93$, $SD = 6.34$; $F(1, 25) = 14.55$, $p < .001$, $\eta_p^2 = .368$). A similar 2 (tracking difficulty) x 2 (priority assignment) repeated measures ANOVA was run to determine the effect of tracking priority on number of switches *to* the tracking task. A main effect of tracking difficulty was found with participants switching to the tracking task more when it was difficult ($M = 24.41$, $SD = 5.83$) than easy ($M = 19.85$, $SD = 6.32$; $F(1, 25) = 13.92$, $p < .001$, $\eta_p^2 = .358$). However, there was no effect of priority $F(1, 25) = .101$, $p = .753$, $\eta_p^2 = .004$, and no interaction was found between difficulty of the tracking task and priority assignment, $F(1, 25)=.101$, $p = .753$, $\eta_p^2 = .004$. Relative to a priority effect, these results are similar to previous work (Gutzwiller, 2014; Wickens et al., 2016).

Visual Dwell and Time to Entry

To test H4, first a more global approach was taken to gauge any relationship between subjectively rated interest of a task and visual dwell time. Specifically, it was hypothesized that the more interesting a participant found a task, the more time a participant would spend visually dwelling on it. Subjective interest ratings of each task were taken from each participant's third (final) questionnaire which, to reiterate, involved paired task comparisons of three attributes relevant to STOM. All participants' ratings of each tasks were then ordered into a ranking system (1 = highest interest, 4 = lowest interest) to establish the most interesting (1) task.

To compare scores with dwell time, the number of participants who ranked a task as most interesting were then compared with an average dwell time per task across those specific participants who rated it as most interesting. For instance, if three participants found the communications task to be most interesting, an average dwell time for communications was taken across those three participants, across both trials. This was done for each task and a Spearman's rank order correlation was then conducted.

Three participants were excluded from the analysis because they gave two tasks the same interest (and highest) interest rating. Initial analysis indicated the relationship was monotonic by assessment of a scatterplot. There was a significant, strong, positive correlation between the number of people rating a task as most interesting and dwelling on it for a cumulatively longer period of time, $r_s(2) = 1.000$, $p < .001$. For a visual representation of the scatter plot, reference Appendix D, Figure 6. This finding aligns with the hypothesis in that the more interesting task (represented by the number 1) has more dwell time associated with it.

However, a narrower, more insightful approach was needed to assess dwell time relationships *within* a task. Consequently, four different Spearman's rank-order correlations were conducted to determine the relationship between subjectively rated interest of each task and the visual dwell time associated with it. Again, the subjective scores were taken from the third of three questionnaires and the task ratings according to the attribute interest were ordered into a ranking system for each participant from most interesting (1) to least interesting (4). The ranked tasks were then compared with the respective task's dwell time, averaged across trials. Three participants were excluded from this analysis because two or more tasks held the same interest rating, leaving it infeasible to be assigned a true rank. This analysis varies from the prior in that all participants (except those excluded) are included, with their respective interest rank for the task and the corresponding dwell times for each task across trials. In contrast, these results found no significant correlations between the tasks ranking and their respective mean dwell times (Monitoring: r_s (24) = -.218, p = .307; Tracking: r_s (24) = -.042, p = .846; Communications: r_s (24) = -.090, p = .676; Resource Management: r_s (24) = -.341, p = .103). For a visual representation of each correlation analysis, reference Appendix D. Again, when dwell times were examined in a more task-centered way, the results show little to no relationship between interest and dwell time. However, when examined in a global sense, a strong, positive relationship was found between the number of participants who found a task to be most interesting and a task's related dwell time. This finding is explored more in the discussion.

Table 3. Subjective Interest ranks and mean dwell times across trials, organized according to task.

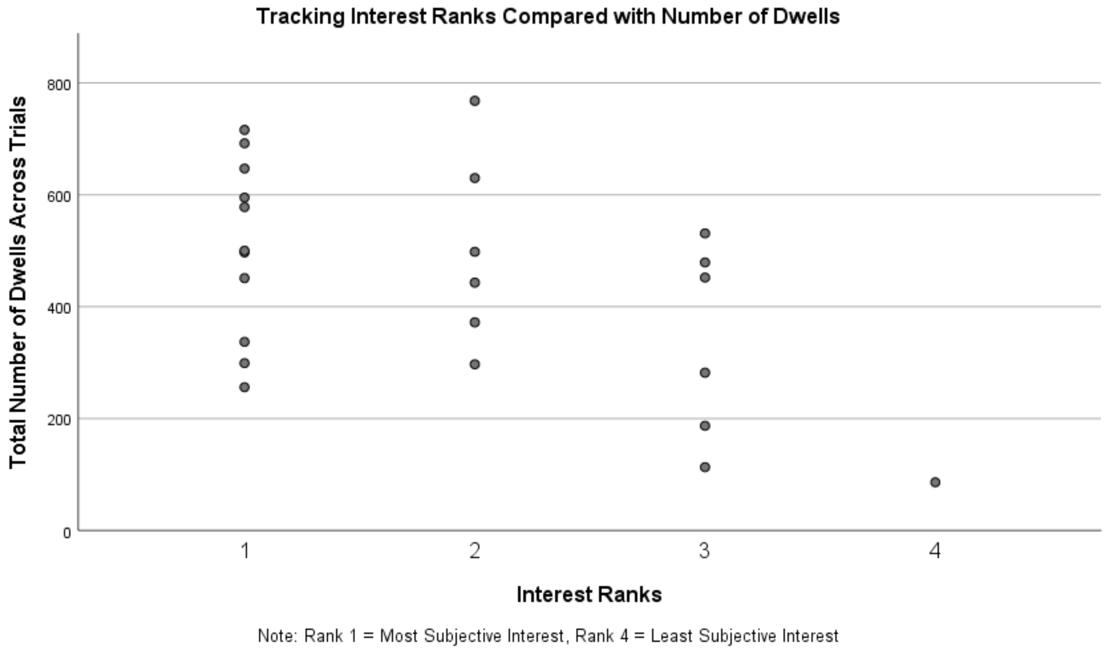
Task	Avg. Rank	Count of Most Interesting Rank	Avg. Dwell Across Trials	Adjusted Dwell
Trk	1.9	11	240.86	242.18
Rman	1.8	10	180.66	186.05
Comm	2.8	3	40.56	44.65
Mon	3.6	0	65.45	0

Note: The **average dwell** times across trials include all participant data whereas, **adjusted dwell** times are calculated by only including the participants who rated that task as most interesting.

Again, it is unknown whether total dwell time for a task is an optimal measure of interest in a task because some tasks simply require more time to complete (e.g., resource management). Instead, for instance, the total number of dwells or the average time spent per dwell within an AOI may provide a clearer picture of how participants are choosing to allocate their time and their attention. Visiting an AOI ten times for ten seconds per visit may reveal a different attention allocation strategy than visiting an AOI 100 times for one second per visit. The former could constitute a more directed, interest-focused strategy or a highly involved task with multiple parts to attend to, whereas the latter could point to a rapid sampling strategy less influenced by such factors and more concerned with attending to all tasks in a coarse-grained approach.

In turn, two separate, exploratory analyses were conducted. First, to see if any relationships existed between ranked tasks according to subjective interest and total amount of dwells across trials, rank correlations were examined. The ranks were taken from the initial analysis conducted for hypothesis four and the same three participants were excluded. Four separate Spearman rank order correlations were run comparing subjectively ranked interest and total amount of dwells per AOI across trials. A significant, moderate, negative correlation was found for the tracking task (r_s (24) = -.432, p = .035) noting that negative correlation actually indicates a positive relationship between interest and amount of dwells. All other analyses returned nonsignificant correlations (Monitoring: r_s (24) = -.051, p = .812; Communications: r_s (24) = -.041, p = .850; Resource Management: r_s (24) = -.090, p = .675). Still, regarding tracking, the higher the interest, the more times the task was visited. Refer to Figure 4 for a visual representation of the correlation found with the tracking task.

Figure 4. The subjective interest ranks for the tracking task compared with its corresponding number of dwells across both trials, per participant.



Second, an additional four Spearman rank order correlations were conducted to assess any relationship between subjectively ranked interest and average time spent per dwell, for each AOI, across both trials. No significant correlations were found (Monitoring: r_s (24) = .266, p = .209; Tracking: r_s (24) = -.256, p = .228; Communications: r_s (24) = .016, p = .940; Resource Management: r_s (24) = -.049, p = .822.

Analysis for hypothesis five also included data derived from eye tracking, however in a different form. To assess whether there was a relationship between subjectively rated priority of a task, and visual time to entry (TTE) of an AOI after an event, ranked tasks were compared on their eye tracking data. An event was considered 1) anything that occurred which required a participant response, or 2) a communication event that did not require subject response, but still sounded. Subjective scores were again taken from the third of the three questionnaire administrations. The task ratings according to the attribute priority were ordered into a ranking system for each participant from highest priority (1) to lowest priority (4). The ranked tasks were then compared with the respective task's time to entry for events occurred (for each participant), averaged across trials.

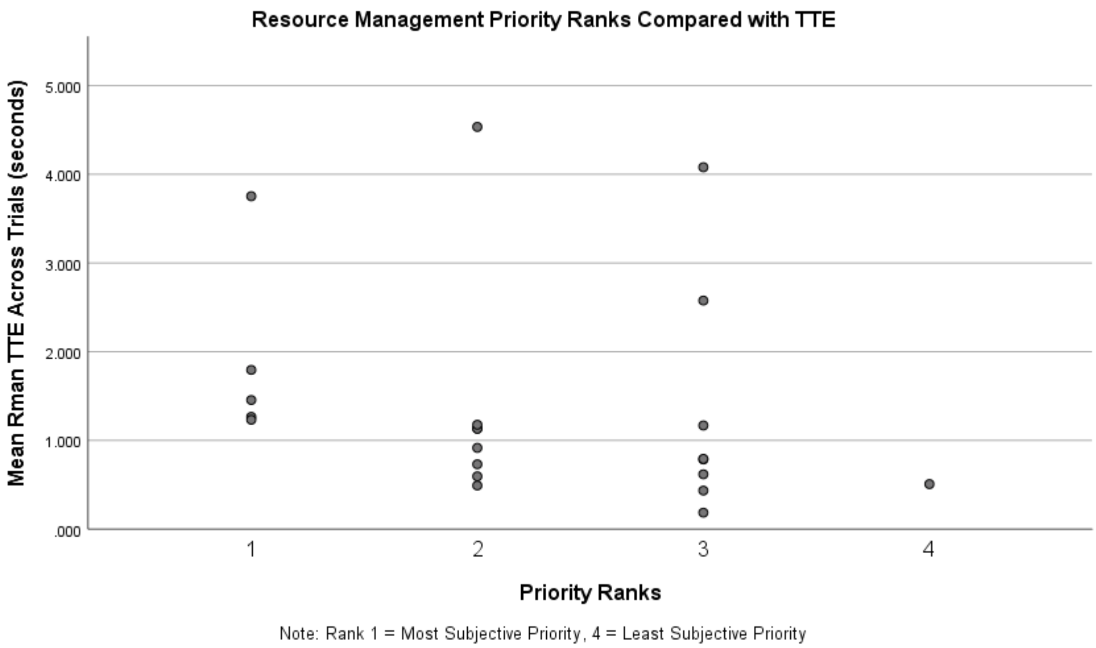
Five participants were excluded from the analysis because they had the same priority score for multiple tasks. Additionally, not all measured events were noticed; within the participants being considered for analysis, 176 events (152 Monitoring events, 23 Communications events, and 1 Resource Management event) across the experiment went unnoticed so there were no time to entry values for them (that is, 15.7% of all events were not noticed, and thus not considered for time to entry analysis).[2] An additional complication was observed; if a participant was already looking in the AOI in which the event occurred, the TTE was scored as zero (there were 209 cases of this occurring, or 18.6% of the total analyzed cases; 74 monitoring events, 81 communication events, and 54 resource management events). The tracking task was not included in the TTE analysis because the nature of the task is continuous and ongoing, with no events to mark the beginning or end of a TTE sequence.

For the tasks Monitoring and Communications, there were no significant correlations (Monitoring: r_s (22) = .069, p = .761; Communications: r_s (22) = .092, p = .683). A moderate, negative correlation was found for Resource Management r_s (22) = -

[2] Again, due to conflict events not occurring simultaneously, the time to entry analysis was only performed on the first event that presented itself in the conflict situations, in addition to all free-standing events. Total events that occurred: 1,670; total events considered for time to entry analysis: 1,122; total events noticed by participants: 946.

.457, *p* =.032 (Figure 5). Again, it is worth noting that these data make up about 66% of informative TTE instances. Nevertheless, these data reveal slower time to entry with higher priority ranks – a finding that is contrary to what was predicted.

Figure 5. The subjective priority ranks for the resource management task compared with its corresponding TTE across both trials, per participant.



Similar to the exploratory analysis of hypothesis four and its justification, it is unclear whether perceived priority of a task influences its TTE after an event occurs. In the case of the resource management task, a better priority rating was moderately correlated with a slower time to entry of the AOI. However, for the other tasks, no such correlation was present. Again, this begs the question of whether the measure of TTE is the optimal way to assess a relationship between subjective priority and attention allocation after an event. So, an additional exploratory set of analyses was conducted to assess whether subjective priority ranks were related to the total number of dwells or average time spent per dwell.

First, four Spearman rank order correlations were run to determine if a relationship existed between total number of dwells across trials and subjective task priority. The tasks' ranks were taken from the initial rank order correlation analysis from hypothesis five and were applied to the following analyses. The same five participants were excluded for having multiple tasks ranked as highest priority. There were no significant correlations found between the two variables (Monitoring: r_s (22) = -.336, p = .127; Tracking: r_s (22) = -.130, p = .563; Communications: r_s (22) = .149, p = .508; Resource Management: r_s (22) = -.124, p = .584).

Finally, four additional Spearman rank order correlations were conducted to reveal any relationship between average time spent per dwell across trials and subjective task priority. No significant correlations were found (Monitoring: r_s (22) = .020, p = .929; Tracking: r_s (22) = -.180, p = .422; Communications: r_s (22) = .073, p = .764; Resource Management: r_s (22) = .046, p = .839). Although the exploratory analyses did not reveal any significant relationships, the lack of findings may also be informative for the overall consideration of using eye tracking to validate elements of STOM. This implication is covered at length in the discussion section.

To assess H6, a 2 (tracking difficulty) x 2 (priority assignment) repeated measures ANOVA was run to determine if the difference in dwell times between high and low priority groups would be greater when the task is more difficult. A significant effect was found for difficulty, $F(1, 23) = 15.72$, $p = .001$, $\eta_p^2 = .406$ with difficult trials resulting in longer dwell times (M = 262.80, SD = 73.78) compared to easy (M = 216.70, SD = 76.13). However, there was no effect of priority condition, $F(1, 23) = 1.49$, $p = .235$, η_p^2 = .061. Although mean differences between the high and low priority groups revealed dwell times of the tracking task to be greater when the task was easy (*MD* = 38.01) than when difficult (*MD* = 29.06), the interaction between difficulty and priority was not significant $F(1, 23) = .150$, $p = .702$, $\eta_p^2 = .006$.

CHAPTER 5

DISCUSSION AND IMPLICATIONS

Overall, the focus of the study was twofold: (1) replicate the experiment from Wickens et al. (2016, Experiment 1) to further validate the STOM model and, (2) to address a limitation of existing STOM validation studies with the addition of eye tracking to act as a proxy for visual attention by means of dwell time and time to entry measures.

The first hypothesis suggested that fewer switches should occur in difficult tracking conditions compared to easy tracking conditions because task switching is believed to be cognitively effortful and resource limited (Wickens et al., 2016). A sensible strategy, indeed, as switching tasks during difficult, overloading conditions could further exacerbate any mental demand and deplete cognitive resources (Wickens & Gutzwiller, 2017). This result was found in prior research (Wickens et al., 2016, experiment 1). However, in the current study this was not the case and, though not significant, fewer switches occurred in the easy conditions. This result could pose contrasting evidence to effort avoidance literature and some previous STOM model validations. One possible explanation for why this may have occurred, aside from the differences between the original study and the current study, is related to the cognitively-based theory proposed by Sheridan (2007). Sheridan proposes that a dynamic AT which is suddenly perturbed would likely prompt a task switch, because additional time spent on the OT could lead to a greater perturbance on ATs. Remembering that the tracking task in the difficult trials did required a greater amount of manual input via the joystick because its reticle updated more often and in a much more dynamic fashion compared to in the easy trials. Therefore, when in this more dynamic state, Sheridan's theory would predict a greater chance of attending to tracking when it was the AT in the difficult condition, counter to the prediction of the STOM model for AT selection. It may have accounted for the greater amount of switches in the difficult trials compared to the easy. Also, perhaps with additional participants and, in turn, an appropriate power, the outcome of the analysis may have fallen in line with previous work.

Second, (H2) it was predicted that a difficult task should invoke fewer switches to it. Being that tracking was the manipulated task, it was assessed for task switches to it. This was found not to be true and in a significant way, as described above; this finding is surprising not only because it opposes the choice probabilities represented in STOM, but also because it resists the attributes, or rational, for switching to an AT. Instead, these results fall in line with Sheridan's theory. But, if humans are considered to conserve cognitive resources and limit high loads, then it would be safe to assume difficult tasks would be avoided and less likely to be switched to as an alternative (Wickens et al., 2015). Generally, a more difficult AT would be considered less attractive, but in the case of this experiment, it was switched to more often than when it was easy. However, once a difficult task is switched to and therefore becomes the OT, the nature of the task is not quite as much of an impediment and there is evidence to suggest a 'stickiness' (Wickens & Gutzwiller, 2017; Wickens et al., 2016, Experiment 2). In other words, a "switching hysteresis" may develop once a participant overcomes the initial effort it takes to switch, and it becomes more likely to stay on the difficult task (p. 3). So, it is important to consider the attribute of difficulty from the perspective of an AT (which has a negative attractiveness weight) and as a part of an OT (which holds a positive attractiveness weight). This is ultimately important because evidence was found in the current study that states the tracking task was switched to and from more often than when it was easy (see analysis of H2 and H3). Yes, by the nature of examining one difficulty-manipulated task, it is probable that if switches to it increase, so too will switches away from it. However, a significant finding of greater switches away from a difficult task directly contradicts the STOM prediction of staying on a difficult task due to switching hysteresis, for example. This finding becomes increasingly complex when dwell times are considered below.

Additionally, although the attribute of difficulty is the main focus here due to it being manipulated, there are other factors that could have influenced task switching behavior. Put simply, it is important to consider other influential STOM factors such as interest, or TOT that may either offset or intensify a bias to switch tasks (Wickens et al., 2015). The co-occurrence of attributes within a task may create tradeoff situations such as in Gutzwiller's research in which a highly salient, low difficulty task was favored over a task that was more interesting and of higher priority (2014). That is also why is it crucial to establish appropriate weights of the attributes so that tradeoffs can be predicted, especially within real-world multitask environments in which the co-occurrence of attributes within tasks is likely. Considering the subjective ratings of interest and difficulty for the tracking task, perhaps in this instance the interest effectively offset the difficulty and produced a noticeable difference in switching to tracking, even when the difficulty is high. This would fall in line with previously mentioned work that found interest to be a more influential attribute of task switching than difficulty (Spink, et al., 2006). However, this could also be argued against and deemed unlikely because switch avoidance is a parameter of STOM with substantial confidence, and a parameter that difficulty is a major part of. Finally, it should be noted that due to lack of participants and therefore appropriate power, the bearing the results hold is limited.

The third hypothesis predicted a main effect for task priority, stating a task with higher priority should reveal fewer switches away from it when it is ongoing, and more switches toward it if it is an alternative. No evidence emerged to suggest an effect of priority exists. That is not to say the priority manipulation was completely unsuccessful. Questionnaire data reveal that the manipulation was effective in shaping subjective prioritization much like in Gutzwiller (Experiment 1, 2014) and Gutzwiller & Sitzman (2017). It is clear that participants understood that tracking was the prioritized task; however, the transfer of that understanding to switch behavior was not as evident. Previous research concerning the influence of priority, although limited, suggests that priority instruction may influence the time spent working on a task but not necessarily the likelihood of switching to it per se (Wickens et al., 2016). Likewise, Raby and Wickens (1994) established pilots too are influenced by priority with regard to the time they allocated to tasks, but not necessarily their task-management decisions (i.e. switches/choices of task).

Eye Tracking Discussion

Upon a more macro-scale analysis of the data, H4 was supported and suggested that more dwell time will be spent on a task when a participant finds the task to be interesting. When considering the 3rd administration of the subjective questionnaires, the tracking and resource management tasks are clearly found be the most interesting across participants (Table 3.). Indeed, those two tasks also received the greatest visual dwell time across trials among participants who rated them as most interesting. However, a remaining question is how much of this dwell time is due to interest rather than the specific nature of the tasks and the attention required to successfully complete them? Both tracking and resource management can be considered continuous tasks in that the main objective of the respective tasks (maintain the reticle centered on the crosshairs; and maintain fuel levels of 2500 units) is ongoing and can necessitate greater dwell times to accomplish as opposed to a task such as monitoring or communications with relatively short, discrete events. As noted, resource management also has discrete events in the form of pump failures that the participants must work around for certain period of time. The act of manipulating the fuel flow and making any necessary changes in combination with continually checking the fuel level status may account for the greater amount of dwell time.

In an effort to explore these results further, additional rank order correlations were run in order to reveal any relationships between a task's interest rank and other visual attention variables. Only one significant result (tracking) suggested the higher the subjective interest, the more dwells the task will incur; this single outcome is consistent with the global results (the more interesting tasks received more overall dwell time). Why this correlation was only found in the tracking task and not the others remains unresolved. In conjunction with the ongoing nature of the task, the priority or subjective difficulty could have made the task more attractive, and thus warrant more dwells. However, when considering the impact of both possible attributes, it becomes unclear what exactly holds more influence than another. For instance, the relative weights of the two attributes could zero out the influence of one another and ultimately have no effect on number of dwells. Future work could specifically focus on the polarization of attributes to see which have more influence. The polarization, in this case, would assess if higher values of the attribute make it more or less attractive.

Given no relationships between time spent per dwell and subjective interest, our analyses did not provide a clearer picture as to how participants allocate their attention, leaving this question unresolved. Overall, there appears to be a global relationship between interest and dwell time, but the exact behavior of that relationship warrants further investigation within each task. One nuance here is that this evaluation took place on only MATB tasks, whereas STOM has utilized other tasks including the BORIS robotic arm, AutoCAMS process control task, and even scheduling software for physicians and questionnaires.

In H5, it was predicted that tasks with higher rated priority should incur lower time to entry to its AOI. Overall, the correlation results showed no relationship for the monitoring or communications task. However, the relationship between resource management and its associated ranks revealed a statistically significant moderate, negative correlation. This finding means as the time to enter an AOI decreased (faster allocation of an event), so too did the subjective priority of the task. Again, for clarity, the ranks are ordered from 1 (highest priority) to 4 (lowest priority). So, the faster the participant entered an AOI post-event, subjective priority of the task was generally lower. This finding is contrary to what was hypothesized and actually reveals slower attention allocation to longer entry times to AOIs after an event is triggered when participants consider the ranked it task as a higher priority. Overall, the results are interesting to consider. Aside from lower power and only 66% of the TTE data being entirely useful, why do the results point towards such contradictions? And why are they only present in the analysis of the resource management task?

One possible answer concerns the nature of TTE analysis and its requirements. The measure is event- and notice-centric, in that it requires events to fire and be noticed to be informative in this context. Evident from the interruption management literature, auditory events are thought to be more salient than visual ones (Lu et al., 2013). Similarly, when the TTE means are examined across trials, separated by tasks, the communication task does boast the lowest (fastest) TTE despite being the task with the least high-priority ranks (For a complete comparison of these means, see Appendix E, Table 8.) Although this difference is not statistically significant, the raw data does support a trend that aligns with previous interruption management literature. The extent that priority has any influence on TTE remains to be seen and, after considering the current results, it most likely does not have any bearing on TTE. One limitation of the current study that is evident here is the lack of TTE measurement for the tracking task. The tracking task was overwhelmingly considered the highest priority due in part to the priority manipulation. However, if future work is to examine this relationship again using similar manipulations, more than likely the use of state space analysis is needed in combination with means squared error (Gutzwiller et al., 2015). Here researchers determined both optimal and non-optimal switching points of the continuous tracking task determined upon the state of the task (stabilized with low error or high error and divergent reticle direction).

Also, mimicking H4, exploratory analyses were conducted to assess any relationships between subjective priority of a task and number of dwells per AOI across trials/average time spent per dwell. No significant results were found when subjective priority was compared to either variable.

The sixth and final hypothesis, which predicted an interaction, was not supported. The difference in tracking task dwell times between high and low priority groups were expected to be greater when the task was more difficult however, there was no significant interaction between the two factors. Instead, more difficult tracking resulted in longer dwell times. This finding is consistent with the *switching hysteresis* hypothesis (Wickens & Gutzwiller 2017) and reveals a positive polarity when a difficult task is the OT. In other words, once a hard task is started, it becomes "sticky" and unsuitable to leave, much like what was found in Gutzwiller, 2014. However, switch analysis in the current study refutes this result (more switches were found from tracking when it was difficult than easy). These two opposing results are even more interesting when dwell time is also considered. Dwell time in the current study stands to support previous findings of the influence of difficulty, but directly contradicts the current study's switch data. This begs the question why the two sources of attention allocation data and task switching are not in alignment and how that affects interpretation in light of previous research. Previous research that also finds evidence of stickiness in a difficult OT has been conducted in environments that are unlike MATB. Finding results in a number of diverse environments, in combination with appropriate power, does increase the chance of finding legitimate results. In a way, this strengthens the results of the eye tracking. The current study did not have adequate power which certainly limits the results of the switching performance.

Concluding Thoughts

To reiterate, multitasking is highly prevalent in real-world occupations. It is in these occupations that humans can emphasize their freedom of choice when tasked with completing several, possibly competing tasks. As technology continues to permeate the work environment and force attention allocation into an indispensable position, task management becomes not only vital but sometimes also a challenge. Considering this, the outcomes of studying attention allocation and task switching could lead to enhanced human-centered, technological design and may provide scientific reason where and when it is needed most.

The current study aimed first to validate the STOM model through replication of prior work (Wickens et al., 2016) and second, to address an existing limitation of validation studies by assessing visual attention switches made by participants as opposed to only performance inputs to determine task switches. Concerning the validation effort, results previously found were not entirely supported. Evidence contradicting the STOM choice probabilities and rational for switching tasks was found. However, there was support for priority neglect (Wickens & Gutzwiller, 2017), in that the priority manipulation did not appear to influence the amount of task switching to and from the tracking tasks. Despite participants acknowledging the priority of the tracking task through subjective questionnaires, it did not seem to alter frequency of task switching. The characterization of priority is also critical. While STOM describes priority as how important a task is relative to an underlying mission and emphasizes any value it may have to overall success, Freed (2000) posits a different view. Instead, the theory argues that reactive prioritization could promote the best of multitask, overload situations through decision-making that uses heuristics and time-relevant information. So, the overall intuitiveness of priority for task attraction is evident.

The inclusion of eye tracking in the validation effort has been productive, despite the lack of amply conclusive results. First, previous findings related to TOT and working memory demand (Gutzwiller, Wickens, & Clegg, 2016) are arguably found in the current study. More complex tasks such as resource management do have a higher dwell time associated with them. Tasks such as resource management or communications could induce high working memory demands they force participants to quickly utilize memory to complete them (recalling failed pumps or frequencies to enter). Consequently, higher dwell times may suggest a strategic task inertia effect, meaning a switch away may exacerbate an already taxed working memory system. This results in greater switch resistance until the task is done.

Additionally, eye tracking data were compared against subjective ratings in hopes of revealing relationships between the two factors that could allow for informed predictions regarding where an operator may dwell or where attention will be allocated first in an overloading situation. Broadly, subjective interest seems to be related to dwell time. As interest in a task rises, so too does the dwell time associated with it. TTE data reveal slower attention allocation to tasks when operators consider the task a higher priority. A significant, moderate correlation was found between the priority of the resource management task and its TTE which reveals slower attention allocation to the task when the priority is higher. Nevertheless, when considering all rank order correlation analyses related to TTE and priority, very little evidence points to the existence of a relationship. Because of this, future work should rather focus on a different attribute, subjectively rated or manipulated, that could be compared with TTE. Salience for instance, although not included in the current version of the Paired Task Questionnaire, could be better suited for TTE analysis because the attribute is more focused on the arrival of an alternative task, similar to the event-centric nature of the TTE measure. Most of all, future work should continue to investigate the influence of difficulty on both an OT and an AT with the assistance of eye tracking. Although switching data in the current study conflicts with previous findings, eye tracking data finds evidence to support and strengthen past work and the STOM model as a whole.

REFERENCES

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science, 16*(6), 351-355. doi:http://dx.doi.org.ezproxy1.lib.asu.edu/10.1111/j.1467-8721.2007.00534.x

Dismukes, R. K. (2010). Remembrance of things future: Prospective memory in laboratory, workplace, and everyday settings. In D. Harris (Ed.), *Reviews of human factors and ergonomics.*

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24*(6), 409-436.

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors, 59*(1), 5-27. doi:http://dx.doi.org.ezproxy1.lib.asu.edu/10.1177/0018720816681350

Freed, M. (2000). Reactive prioritization. *Proceedings NASA International Workshop on Planning and Scheduling in Space*. Washington, DC: NASA.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics*, *24*(6), 631–645. https://doi.org/10.1016/S0169-8141(98)00068-7

Gutzwiller, R. S. (2014). Switch choice in applied multi-task management. *Doctoral Dissertation*, 1–179.

Gutzwiller, R. S., & Sitzman, D. M. (2017). Examining task priority effects in multi-task management. *Proceedings of the Human Factors and Ergonomics Society*, 762–766. https://doi.org/10.1177/1541931213601675

Gutzwiller, R. S., Wickens, C. D., & Clegg, B. A. (2014). Workload overload modeling: An experiment with MATB II to inform a computational model of task management. *Proceedings of the Human Factors and Ergonomics Society*, 849–853. https://doi.org/10.1177/1541931214581179

Gutzwiller, R. S., Wickens, C. D., & Clegg, B. A. (2015). The role of individual differences in executive attentional networks and switching choices in multi-task management. *Proceedings of the Human Factors and Ergonomics Society*, 632–636. https://doi.org/10.1177/1541931215591138

Gutzwiller, R. S., Wickens, C. D., & Clegg, B. A. (2016a). The role of time on task in multi-task management. *Journal of Applied Research in Memory and Cognition*, *5*(2), 176–184. https://doi.org/10.1016/j.jarmac.2016.04.003

Gutzwiller, R. S., Wickens, C. D., & Clegg, B. A. (2019). The role of reward and effort over time in task switching. *Theoretical Issues in Ergonomics Science, 20*(2), 196–214. https://doi.org/10.1080/1463922X.2018.1522556

Holmqvist, K., Nystrom, M., Andersson, R., Dewhurst, R., Jarodzka, H., van de Weiger, J. (2011). *Eye tracking A comprehensive guide to methods and measures.* Oxford, U.K.: Oxford University Press.

Janssen, C. P., Brumby, D. P., & Garnett, R. (2012). Natural break points: the influence of priorities and cognitive and motor cues on dual-task interleaving. *Journal of Cognitive Engineering and Decision Making*, *6*(1), 5–29. https://doi.org/10.1177/1555343411432339

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, *36*(6), 661–679. https://doi.org/10.1017/S0140525X12003196

Lin, Y., & Zhang, W. J. (2003). Evaluating interface usability based on eye movement and hand movement behavioral parameters. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *47*(3), 653–657. https://doi.org/10.1177/154193120304700384

Loukopoulos, L., Dismukes, R., & Barshi, I. (2009). *The multitasking myth*. Farnham: Ashgate Publishing, Ltd.

Lu, S. A., Wickens, C. D., Prinet, J. C., Hutchins, S. D., Sarter, N., & Sebok, A. (2013). Supporting interruption management and multimodal interface design: Three meta-analyses of task performance as a function of interrupting task modality. *Human Factors*, *55*(4), 697-724. doi:http://dx.doi.org.ezproxy1.lib.asu.edu/10.1177/0018720813476298

Meyer, D. E., Kieras, D. E., Allard, T., Chipman, S., Hawkins, H., Vaughan, W., … Jones, C. (1997). A computational theory of executive cognitive processes and multiple-task performance: part 1. basic mechanisms. *Psychological Review*, *104*(1), 3–65. https://doi.org/10.1037/0033-295X.104.1.3

Nothdurft, H. (2000). Salience from feature contrast: Additivity across dimensions. *Vision Research*, *40*(10-12), 1183-1201. https://doi.org/10.1016/S0042-6989(00)00031-6F

Raby, M., & Wickens, C. D. (1994). Strategic workload management and decision biases in aviation. *International Journal of Aviation Psychology*, *4*, 211-240.

Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock, J. R. (2011). *The Multi-*

Attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research : A User's Guide.

Senders, J. W. (1983). *Visual scanning processes*. Tilburg: University of Tilburg Press.
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

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, *115*(1), 101-130. doi:http://dx.doi.org.ezproxy1.lib.asu.edu/10.1037/0033-295X.115.1.101

Salvucci, D.D., & Taatgen, N.A. (2011). *The multi-tasking mind.* Oxford, U.K.: Oxford University Press.

Salvucci, D. D., Taatgen, N. A., & Kushleyeva, Y. (2006). Learning when to switch tasks in a dynamic multitasking environment. *Proceedings of the Seventh International Conference on Cognitive Modeling*, 1–6.

Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock, J. R. (2011). *The Multi-Attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research: A User's Guide*. *NASA Technical Memorandum 2011-217164*.

Sebok, A. (2017). The strategic task overload model: History, status, challenges, and extensions into new domains. *Proceedings of the Human Factors and Ergonomics Society*, 755–756. https://doi.org/10.1177/1541931213601673

Senders, J. W. (1983). *Visual Sampling Processes*. Hillsdale: Erlbaum

Sheridan, T. (2007). Attention and it's allocation: Fragements of a model. In A. Kramer, D. A. Wiegmann, & A. Kirlik (Eds.), *Attention: From Theory to Practice* (pp. 16-26). Oxfod University Press.

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience, 6*(12), 1317-1322.

Spink, A., Park, M., and Koshman, S. (2006). Factors affecting assigned information problem ordering during Web search: An exploratory study. Information Processing & Management, 42, 1366–1378. https://doi.org/10.1016/j.ipm.2006.01.007

Trafton, J. G., & Monk, C. A. (2007). Task Interruptions. *Reviews of Human Factors and Ergonomics*, *3*(1), 111–126. https://doi.org/10.1518/155723408x299852

Wickens, C. D. (2015). Noticing events in the visual workplace: The SEEV and NSEEV models. *The Cambridge Handbook of Applied Perception Research*, 749–768.

https://doi.org/10.1017/CBO9780511973017.046

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, *50*(3), 449–455. https://doi.org/10.1518/001872008X288394

Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science*, *3*(2), 159-177.

Wickens, C. D., & Gutzwiller, R. S. (2017). The status of the strategic task overload model (STOM) for predicting multi-task management. *Proceedings of the Human Factors and Ergonomics Society*, 757–761.https://doi.org/10.1177/1541931213601674

Wickens, C. D., Gutzwiller, R. S., & Santamaria, A. (2015). Discrete task switching in overload: A meta-analyses and a model. *International Journal of Human Computer Studies*, *79*, 79–84. https://doi.org/10.1016/j.ijhcs.2015.01.002

Wickens, C. D., Gutzwiller, R. S., Vieane, A., Clegg, B. A., Sebok, A., & Janes, J. (2016). Time Sharing between robotics and process control: validating a model of attention switching. *Human Factors*, *58*(2), 322–343. https://doi.org/10.1177/0018720815622761

Wickens, C. D., Laux, L., Hutchins, S., & Sebok, A. (2014). Effects of sleep restriction, sleep inertia, and overload on complex cognitive performance before and after workload transition: A meta analysis and two models. *Proceedings of the Human Factors and Ergonomics Society*, 839–843. https://doi.org/10.1177/1541931214581177

Wickens, C. D., Santamaria, A., & Sebok, A. (2013). A computational model of task overload management and task switching. *Proceedings of the Human Factors and Ergonomics Society*, 763–767. https://doi.org/10.1177/1541931213571167

APPENDIX A

PAIRED TASK QUESTIONNAIRE EXAMPLE

Paired Task Survey 1

Based on the trial you just completed, you should compare several tasks to each other by rating which task represents the category more. For example, you might be asked to compare two tasks (A and B) on which was a higher priority to perform.

Task A ←*--higher task priority--lower -- 0 (equal) -- lower --higher task priority*→Task B

So, if you thought Task A was *higher priority* than Task B, then you would circle the "3" closest to Task A; if however, you thought they were equal, you would circle "0".

COMPARE TASKS FOR PRIORITY – WHICH TASK WAS HIGHER PRIORITY?

Monitoring Task ---- 3 -------- 2 -------- 1 ----0---- 1 -------- 2 -------- 3 ---- Resource Management

Tracking Task ---- 3 -------- 2 -------- 1 ----0---- 1 -------- 2 -------- 3 ---- Resource Management

Communication Task ---- 3 -------- 2 -------- 1 ----0---- 1 -------- 2 -------- 3 ---- Resource Management

Monitoring Task ---- 3 -------- 2 -------- 1 ----0---- 1 -------- 2 -------- 3 ---- Tracking Task

Monitoring Task ---- 3 -------- 2 -------- 1 ----0---- 1 -------- 2 -------- 3 ---- Communications Task

Tracking Task ---- 3 -------- 2 -------- 1 ----0---- 1 -------- 2 -------- 3 ---- Communications Task

COMPARE TASKS FOR INTEREST – WHICH TASK WAS MORE INTERESTING?

Monitoring Task ---- 3 -------- 2 -------- 1 ----0---- 1 -------- 2 -------- 3 ---- Tracking Task

Monitoring Task ---- 3 -------- 2 -------- 1 ----0---- 1 -------- 2 -------- 3 ---- Communications Task

Tracking Task ---- 3 -------- 2 -------- 1 ----0---- 1 -------- 2 -------- 3 ---- Communications Task

Monitoring Task ---- 3 -------- 2 -------- 1 ----0---- 1 -------- 2 -------- 3 ---- Resource Management

Tracking Task ---- 3 -------- 2 -------- 1 ----0---- 1 -------- 2 -------- 3 ---- Resource Management

Communication Task ---- 3 -------- 2 -------- 1 ----0---- 1 -------- 2 -------- 3 ---- Resource Management

COMPARE TASKS FOR DIFFICULTY – WHICH TASK WAS MORE DIFFICULT?

Monitoring Task ---- 3 -------- 2 -------- 1 ----0---- 1 -------- 2 -------- 3 ---- Resource Management

Monitoring Task ---- 3 -------- 2 -------- 1 ----0---- 1 -------- 2 -------- 3 ---- Tracking Task

Monitoring Task ---- 3 -------- 2 -------- 1 ----0---- 1 -------- 2 -------- 3 ---- Communications Task

Tracking Task ---- 3 -------- 2 -------- 1 ----0---- 1 -------- 2 -------- 3 ---- Communications Task

Tracking Task ---- 3 -------- 2 -------- 1 ----0---- 1 -------- 2 -------- 3 ---- Resource Management

Communication Task ---- 3 -------- 2 -------- 1 ----0---- 1 -------- 2 -------- 3 ---- Resource Management

APPENDIX B

PERFORMANCE DESCRIPTIVES

Table 4.
Performance Descriptives A.
The mean performance of each task divided by the between subjects condition of priority assignment. Tasks are reported in terms of overall accuracy.

Condition/Difficulty	Monitoring F5 Acc.	Monitoring F6 Acc.	Monitoring Scales Acc.
Equal Priority			
Easy Trk	55%	86%	41%
Difficult Trk	43%	88%	55%
Tacking Priority			
Easy Trk	42%	90%	32%
Difficult Trk	33%	94%	45%

Table 5.
Performance Descriptives B.
The mean performance of each task divided by the between subjects condition of priority assignment. The Comm task is reported in terms of overall accuracy and the Tracking and Resource Management task is reported in root mean squared deviations.

	Comm. Radio Acc.	Comm. Frequency Acc.	Tracking Error	Rman Error
Equal Priority				
Easy Trk	60%	56%	21.4	23.4
Difficult Trk	88%	67%	43.6	8.7
Tacking Priority				
Easy Trk	68%	64%	17.09	-147.26
Difficult Trk	73%	64%	36.07	-100.07

APPENDIX C

SUBJECTIVE RATINGS DESCRIPTIVES

Table 6.
Subjective Ratings Descriptives.
The mean subjective ratings for each task according to each STOM attribute across both conditions. Higher ratings reflect a better overall global attractiveness for the attributes Priority and Interest. Lower ratings are better for Difficulty.

Attribute	Task	Statistic	T1	T2	T3
Priority					
	Mon	Mean	-3.37	-2.96	-3.52
		SD	3.05	3.77	3.89
	Trk	Mean	3.04	3.78	3.89
		SD	4.26	3.59	3.77
	Comm	Mean	-0.52	-2.78	-2.26
		SD	5.82	4.74	3.87
	Rman	Mean	0.85	1.96	1.89
		SD	3.40	3.35	3.53
Interest					
	Mon	Mean	-4.11	-4.78	-4.52
		SD	3.62	3.37	2.74
	Trk	Mean	3.85	2.22	2.52
		SD	3.05	3.95	4.28
	Comm	Mean	0.48	-0.52	-1.33
		SD	4.75	5.03	4.99
	Rman	Mean	-0.22	3.07	3.33
		SD	3.07	3.52	3.22
Difficulty					
	Mon	Mean	-3.78	-3.26	-1.63
		SD	3.64	4.10	4.29
	Trk	Mean	2.41	1.22	1.11
		SD	3.98	4.46	3.71
	Comm	Mean	-1.33	-1.37	-2.89
		SD	4.17	5.20	4.21
	Rman	Mean	2.70	3.41	3.41
		SD	3.54	3.73	3.58

Table 7.
Task-Focused Subjective Ratings.
The ratings represent the mean value across both conditions and only 2nd and 3rd trials except for the priority rating of the tracking task which is separated. Higher ratings reflect a better overall global attractiveness for the attributes Priority and Interest. Lower ratings are better for Difficulty.

Task/Attribute	Priority	Interest	Difficulty
Tracking		2.4	1.2
Equal	2.3		
Tracking	5.5		
Monitoring	-3.2	-4.6	-2.4
Communications	-2.5	-0.9	-2.1
Resource Management	1.9	3.2	3.4

APPENDIX D

SPEARMAN RANK ORDER SCATTER PLOTS

Figure 6. Global Dwell Time & Highest Rank Instances.
Global assessment of dwell time and subjectively rated interest. The number of participants who rated a task as most interesting was compared with the dwell time for that said, averaged across those participants.

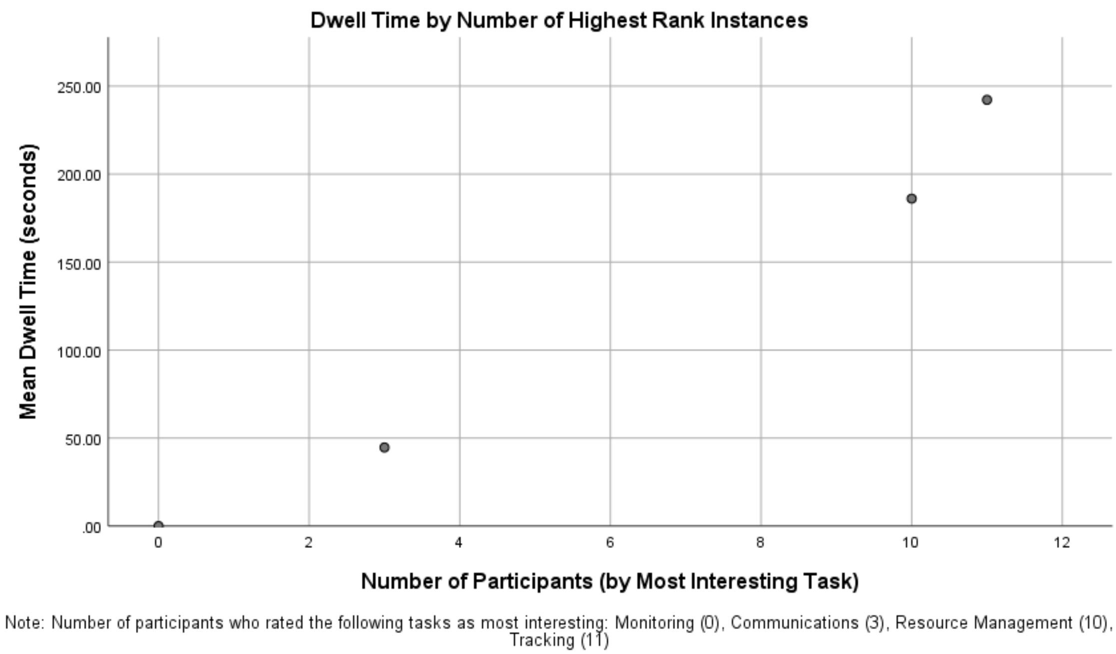


Figure 7. Monitoirng Dwell Time & Ranks of Interest.
The subjective interest ranks for the tracking task compared with its corresponding dwell time across both trials, per participant.

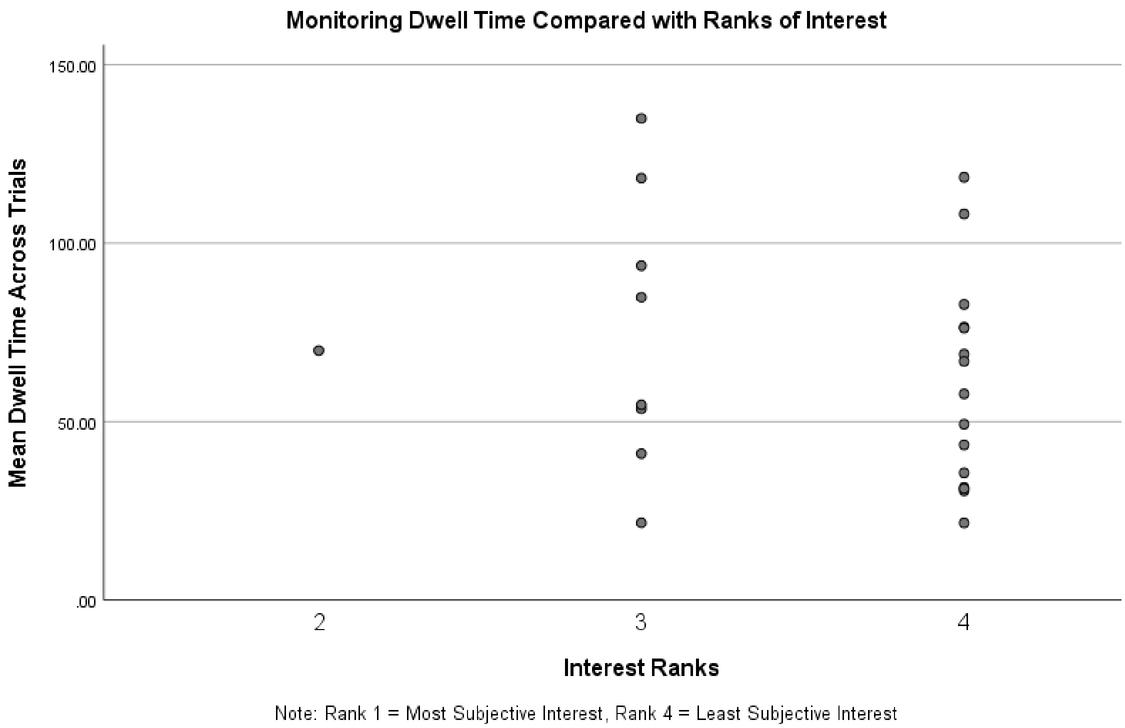


Figure 8. Tracking Dwell Time & Ranks of Interest.
The subjective interest ranks for the tracking task compared with its corresponding dwell time across both trials, per participant.

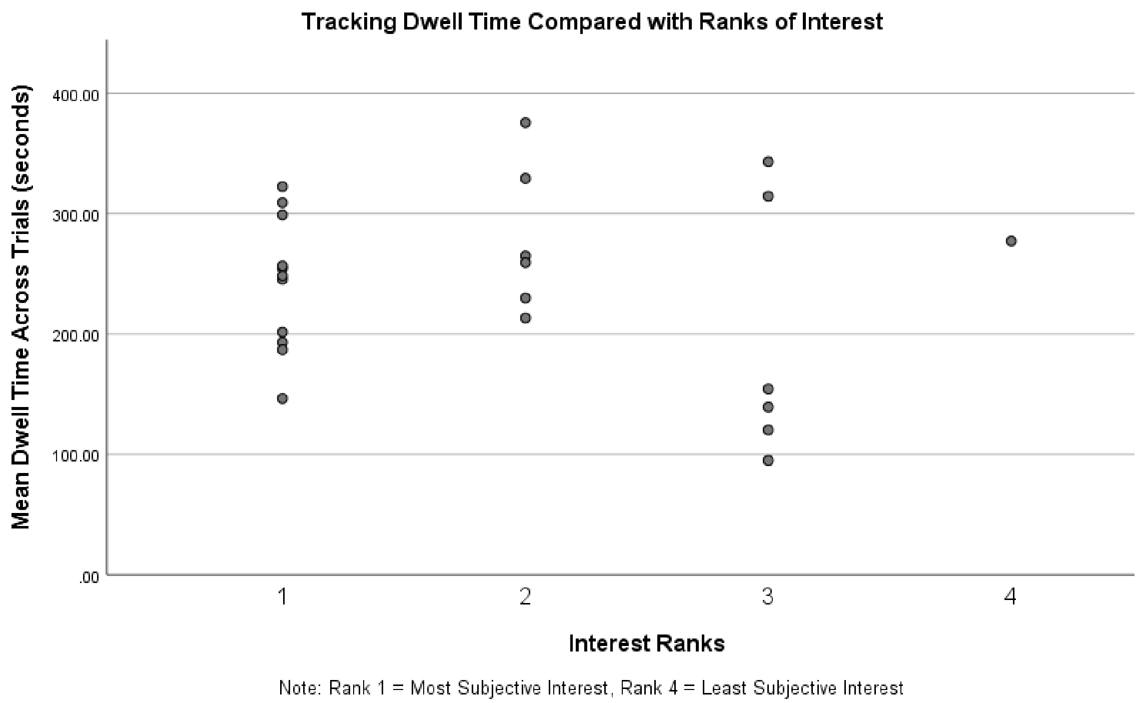


Figure 9. Communications Dwell Time & Ranks of Interest.
The subjective interest ranks for the communication task compared with its corresponding dwell time across both trials, per participant.

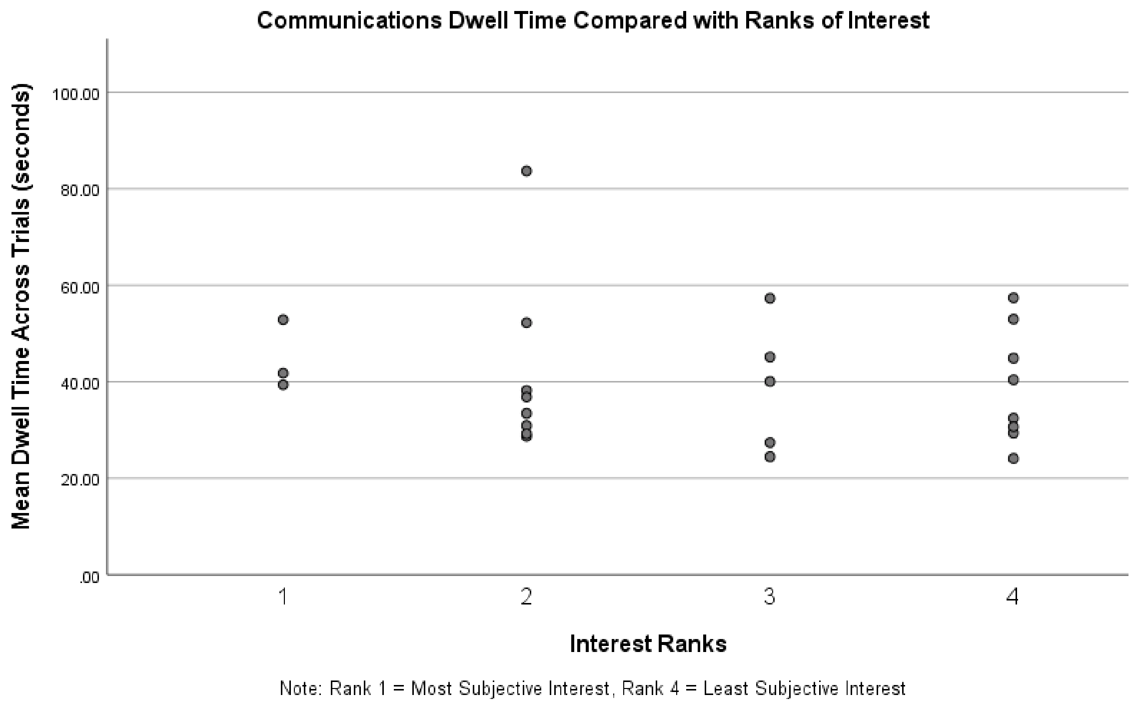
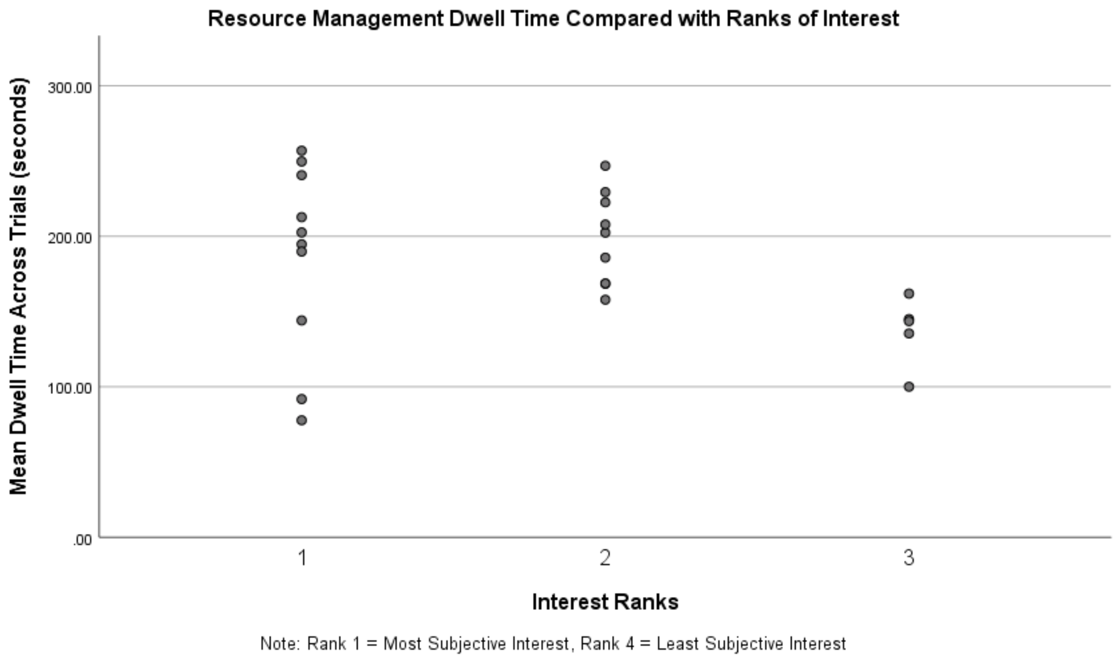


Figure 10. Resource Management Dwell Time & Ranks of Interest.
The subjective interest ranks for the resource mangement task compared with its corresponding dwell time across both trials, per participant.



Priority Rank & Time To Entry Scatter Plots

Figure 11. Monitoring TTE & Ranks of Priority.
The subjective priority ranks for the monitoring task compared with its corresponding TTE across both trials, per participant.

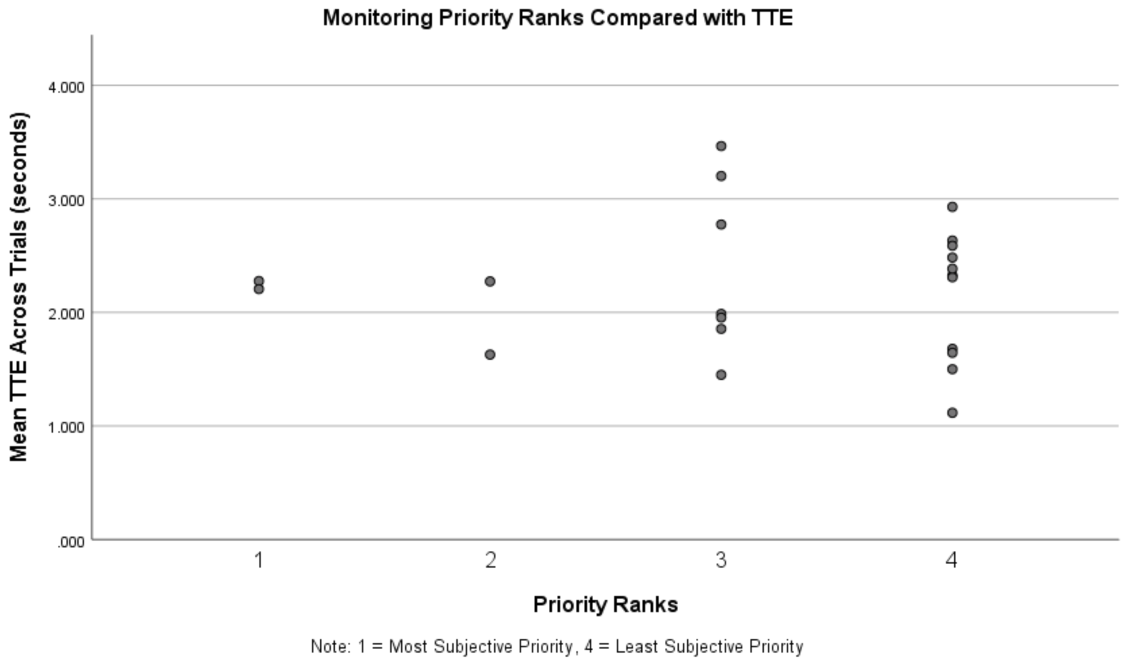
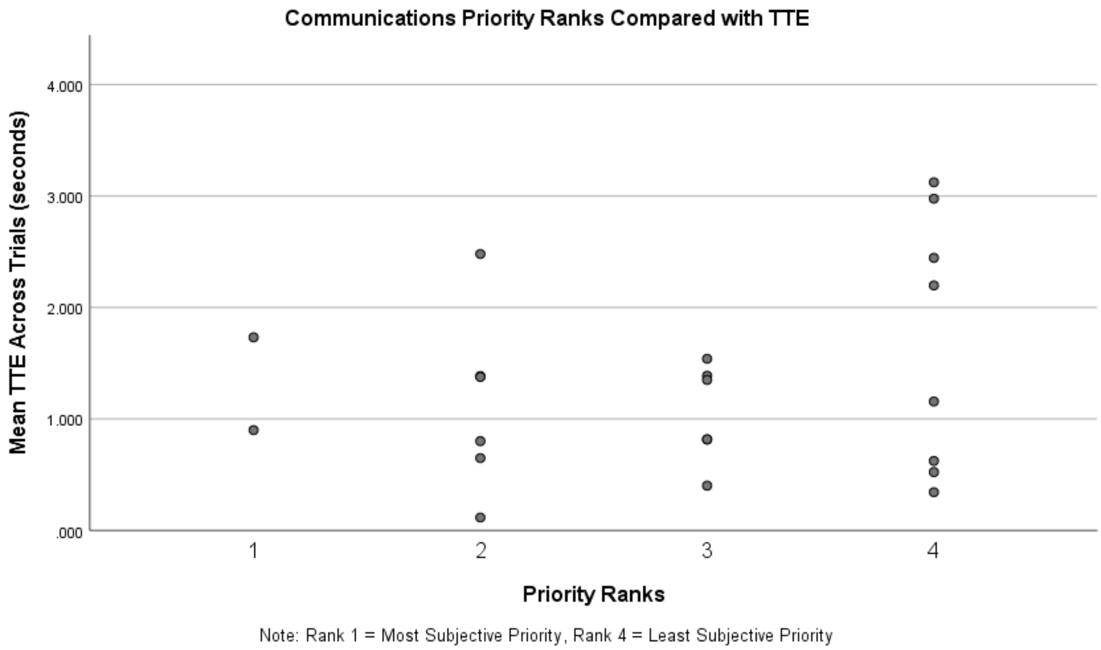


Figure 12. Communications TTE & Ranks of Priority.
The subjective priority ranks for the communications task compared with its corresponding TTE across both trials, per participant.



APPENDIX E

EYE TRACKING DATA TABLES

Table 8.
Priority Ranks & Mean TTE.
The count of highest subjective priority rank listed with its corresponding time to entry (averaged across trials), all separated by task.

Note: The average TTE times across trials include all participant data whereas, adjusted TTE times are calculated by only including the participants who rated that task as highest priority.

Task	Avg. Rank	Count of Highest Priority Rank	Avg. TTE Across Trials	Adjusted TTE
Trk	1.6	13		
Rman	2.2	5	1.425	1.9
Mon	3.2	2	2.212	2.241
Comm	2.9	2	1.325	1.316

Table 9.
Mean Total Dwell Time Separated by Condition.
The mean total dwell time (seconds) on each task in the MATB II organized by the between subjects condition of priority assignment.

Condition/Difficulty	Monitoring	Tracking	Communications	Resource Management
Equal Priority				
Easy Trk	83.28	198.44	37.24	225.63
Difficult Trk	61.38	248.85	32.06	188.62
Tracking Priority				
Easy Trk	68.59	236.48	45.36	181.18
Difficult Trk	49.10	277.91	48.91	151.07

APPENDIX F

EXPLORATORY ANALYSIS MEANS

Table 10.
Mean Number of Dwells.
The mean total number of dwells of each task organized by the between subjects condition of priority assignment.

Condition/Difficulty	Monitoring	Tracking	Communications	Resource Management
Equal Priority				
Easy Trk	120	220	50	162
Difficult Trk	108	273	47	181
Tracking Priority				
Easy Trk	95	202	67	161
Difficult Trk	74	232	56	160

Table 11.
Average Time Spent Per Dwell.
The average time spent per dwell (seconds) on each task in the MATB II organized by the between subjects condition of priority assignment.

Condition/Difficulty	Monitoring	Tracking	Communications	Resource Management
Equal Priority				
Easy Trk	1.58	1.13	1.37	0.75
Difficult Trk	1.68	1.11	1.58	1.13
Tracking Priority				
Easy Trk	1.61	0.99	1.53	0.92
Difficult Trk	1.63	0.87	1.21	1.08

APPENDIX G

MATB INSTRUCTIONAL POWERPOINT

Learning how to perform MATB II

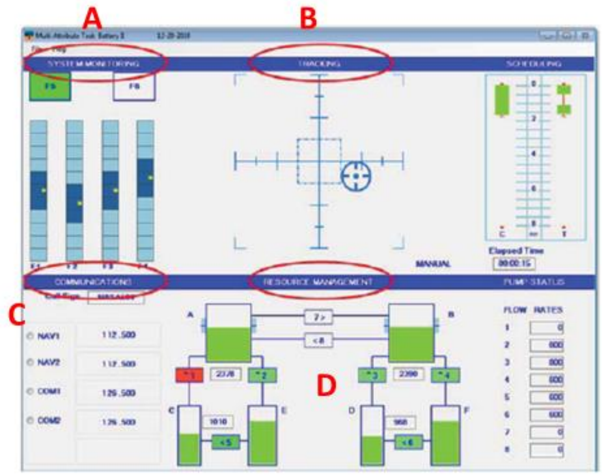
Welcome

- Please take 10 minutes to review the following materials. You can return to previous material to review as often as you'd like after you've reviewed the entire document once, up until time is called.
- At the end of the time, you'll be asked to actually perform these tasks in the simulation – so pay attention! No one can help you after the training period is over.

This is an overview of the main display that you will see during the experiment.

The main display shows the inner workings of several tasks (shown here marked A, B, C, and D).

Performance of these tasks is done simultaneously, but you will learn about each system one by one.



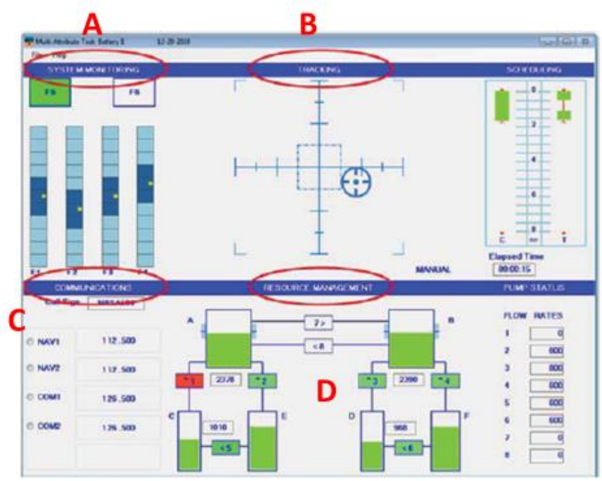
Four tasks simulate the kinds of tasks that REAL pilots perform during flight of a commercial jetliner.

These tasks are *critical* to master, as you will be asked to perform all of them.

These tasks include:

a) monitoring,
b) tracking,
c) communicating,
d) and resource management

Each of these tasks has been labeled – take a moment to find them on the main display



System Monitoring Task (task A)

The monitoring task appears in the upper left of the main display. There are 2 main components: At the top, two squares are shown. The rest of the task shows 4 sliding scales, which will be explained shortly.

The left square at the top should appear green; however, it may appear white if it is turned off. *You should respond by clicking it to turn the button back on.*

The right square at the top should appear white; however, it may appear red indicating a failure. *You should respond by clicking the button to turn it off.*

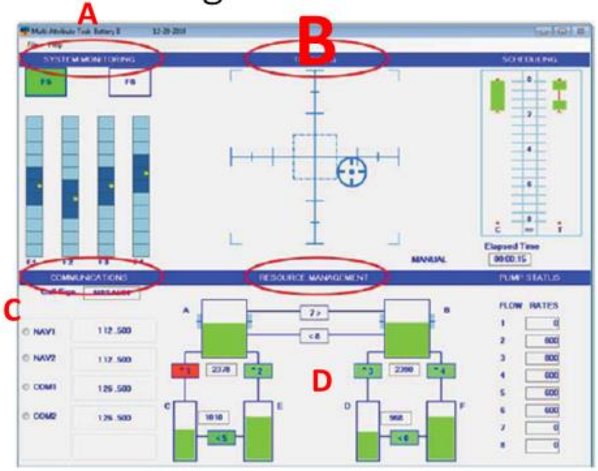
System Monitoring Task (task A)

The lower portion shows 4 sliding scales, each of which has an indicator.

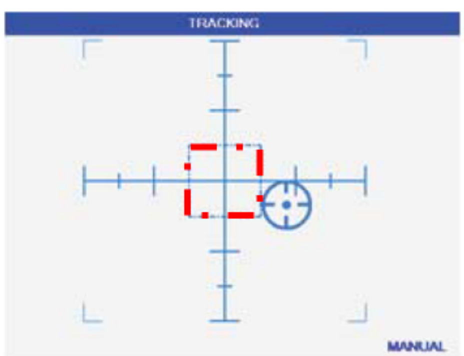
Each indicator can go "out of range" on each sliding scale, either too high or too low. **Your job is to detect these deviations, and click on the scale with the deviation when it occurs.**

Remember, you are trying to quickly, and accurately, detect and respond to these events.

Task B = Tracking Task



Tracking Task (Task B)

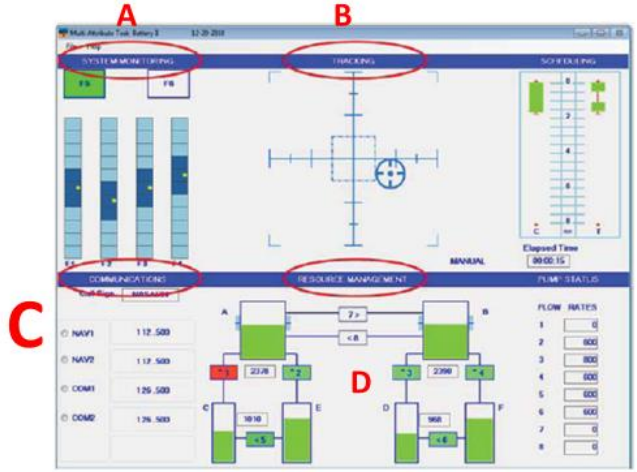


For this task, you will use the joystick to control the small, circular target reticle. Your job is to keep the circular target WITHIN THE BOUNDARIES of the smaller square and as much on the cross as possible.

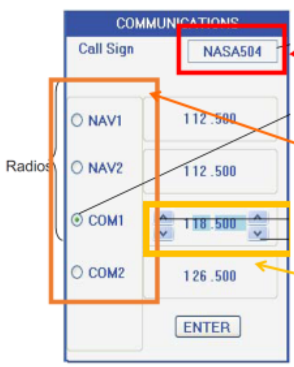
You can think of this like piloting a plane under turbulence; heavy winds will alter the aircraft's path – your job is to keep on course.

If you find the circular reticle drifting out of the center, quickly use the joystick to re-center it. Push the joystick forward to move the reticle up and pull backward to move down.

Task C = Communications Task



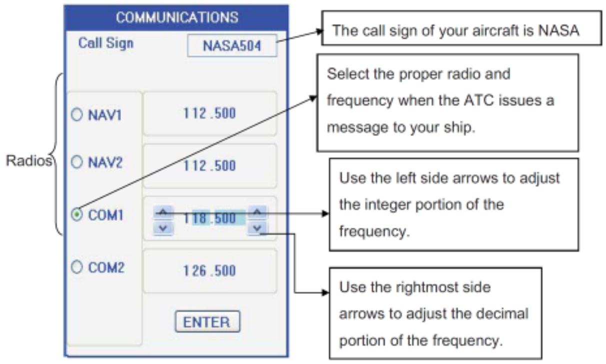
Communication Task (Task C)



You will be hearing messages through your headphones that call a particular aircraft by a "call sign" assigned to it. The call sign will be spoken twice followed by instructions.

When you receive a message it will contain two parts; the first part is the call sign. The second part is a command to select & tune one of the 4 radios to a particular frequency.

Communication Task (Task C)

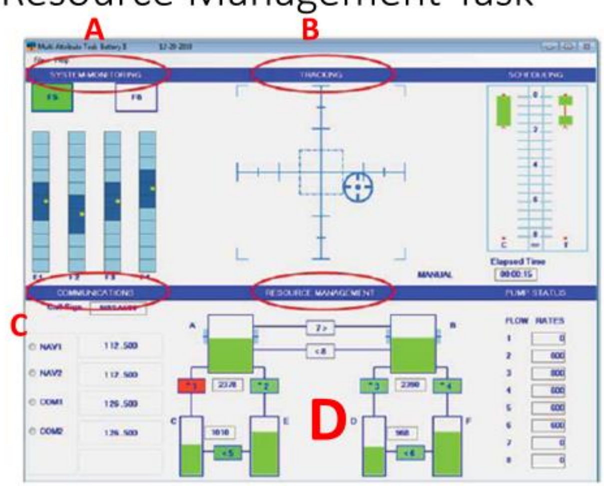


The overall goals are:

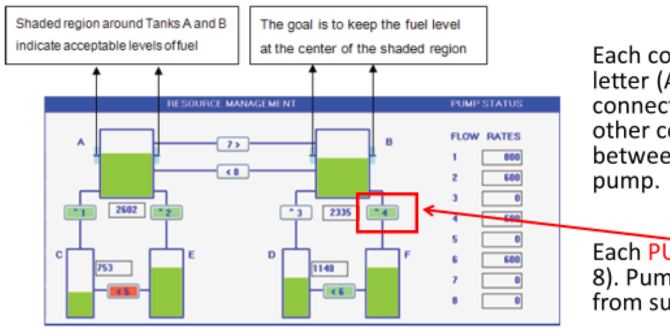
1) Determine whether the message is intended for you: does it match your call sign?

2) IF IT MATCHES, quickly and accurately select the radio, and frequency and adjust it according to the instructions.

Task D: Resource Management Task



Resource Management Task (Task D)

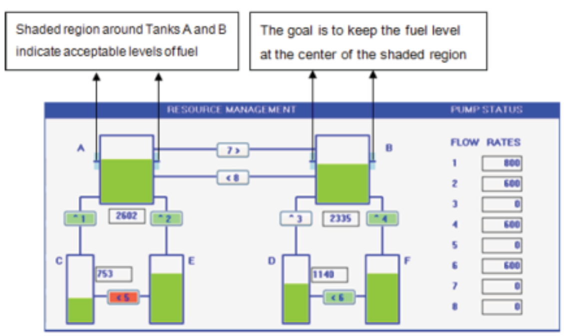


The resource management task represents the necessity to manage fuel aboard an aircraft (shown as green, within the containers).

Each container is designated with a letter (A through F). Pipes (lines) connect certain containers with other containers, and the flow of fuel between them is controlled by a pump.

Each PUMP is numbered (1 through 8). Pumps are used to transfer fuel from supply tanks to the main tanks.

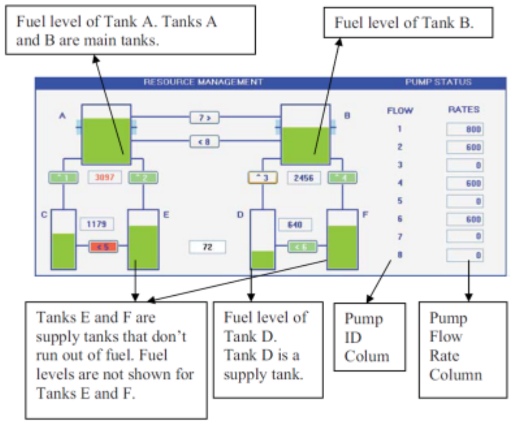
Resource Management Task (Task D)



Pumps are color-coded:

- **green**: pump is ON and flowing in the direction of it's arrow; can be turned off by clicking on it
- White: pump is OFF, but can be turned on if needed by clicking on it
- **Red**: pump has FAILED, BUT you MUST continue to keep the center fuel levels at the center of the shaded region. Try using other pumps.

Resource Management Task (Task D)

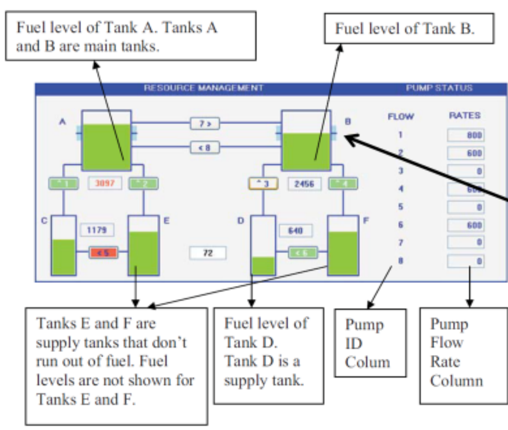


The numbers underneath the two main containers represent the amount of fuel in the container.

Tanks C and D are supply tanks, and contain a maximum of 2000 units. Tanks E and F are unlimited.

You must transfer fuel **to** tanks A and B, via transfers from C and D. This is because tanks A and B continually use fuel to keep the engine going.

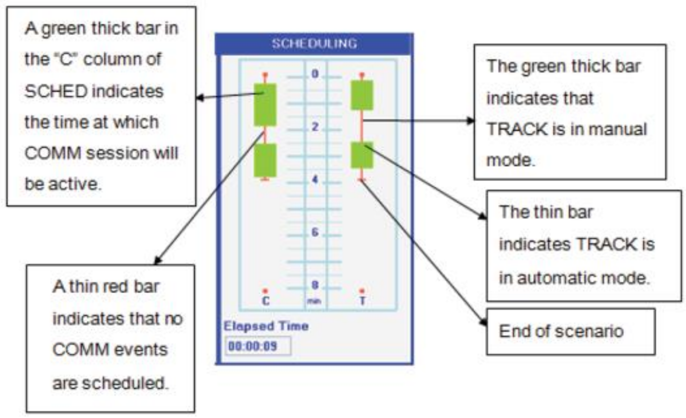
Resource Management Task (Task D)



You cannot let containers A, B, C, or D hit empty. They will become unable to accept transfers of fuel at that point.

The goal of the task is to keep containers A and B filled to within the range indicated by the shaded regions (2500 units in each container).

Task Scheduler Display (upper right corner)



Task Scheduler Display (upper right corner)

- As shown, the task scheduler may help you allocate your attention to upcoming tasks (but it only shows the tracking task, and the communications task).
- It does not show the resource management events, or the monitoring task events.

Final Performance:

- General goals:
  - You will be performing all of the tasks by switching between them.
  - Your goal is to be as fast and as accurate as possible and follow any additional instructions or priorities.
- You will complete a basic training trial next to increase your familiarity.
- Feel free to return to earlier slides if the time has not run out.
- Otherwise wait for your next instructions.