

**MODIFIED COOPER HARPER SCALES
FOR ASSESSING UNMANNED VEHICLE DISPLAYS**

Birsen Donmez
A.S. Brzezinski
Hudson Graham
M.L. Cummings

MASSACHUSETTS INSTITUTE OF TECHNOLOGY*
PREPARED FOR US ARMY ABERDEEN TESTING CENTER

HAL2008-06

DECEMBER 2008



*MIT Department of Aeronautics and Astronautics, Cambridge, MA 02139

Table of Contents

LIST OF FIGURES.....	III
LIST OF TABLES.....	IV
ABSTRACT.....	1
INTRODUCTION.....	2
BACKGROUND	4
<i>Cooper-Harper Scale for Aircraft Handling Qualities.....</i>	<i>4</i>
<i>Modified Cooper-Harper for Unmanned Vehicle Displays</i>	<i>5</i>
GENERAL MCH-UVD DIAGNOSIS TOOL.....	8
<i>Decision Diamond #1: Cannot Complete Mission Task Safely.....</i>	<i>9</i>
<i>Decision Diamond #2: Display Does Not Aid Mission Task Completion.....</i>	<i>10</i>
<i>Decision Diamond #3: Display Does Not Promote Effective Mission Re-planning.....</i>	<i>12</i>
<i>Acceptable Displays</i>	<i>12</i>
<i>Summary of the General MCH-UVD Diagnosis tool</i>	<i>13</i>
GENERATING SPECIFIC MCH-UVDS.....	14
<i>Unmanned Vehicle Mission Taxonomies.....</i>	<i>14</i>
<i>Guiding Principles of Specific MCH-UVD Tool Generation.....</i>	<i>17</i>
<i>Decision Diamond #1: Cannot Complete Mission Task Safely.....</i>	<i>17</i>
<i>Decision Diamond #2: Display Does Not Aid Mission Task Completion.....</i>	<i>18</i>
<i>Decision Diamond #3: Does Not Promote Effective Mission Re-planning.....</i>	<i>20</i>
<i>Acceptable Displays</i>	<i>22</i>
<i>Example Specific MCH-UVD Scale for UGV missions</i>	<i>22</i>
VALIDATION EXPERIMENT.....	24
<i>Participants</i>	<i>24</i>
<i>Apparatus</i>	<i>24</i>
<i>Experimental Design.....</i>	<i>26</i>
<i>Experimental Tasks</i>	<i>26</i>
<i>Dependent Variables.....</i>	<i>29</i>
<i>Procedure</i>	<i>29</i>
EXPERIMENTAL RESULTS	30
<i>MCH-UVD Ratings</i>	<i>30</i>
<i>Display Deficiencies Identified</i>	<i>31</i>
<i>Feedback on MCH-UVD.....</i>	<i>34</i>
DISCUSSION	35
<i>How to Administer MCH-UVD.....</i>	<i>37</i>
CONCLUSION.....	37
ACKNOWLEDGEMENTS.....	37
REFERENCES.....	38
APPENDICES	40

LIST OF FIGURES

Figure 1. Original Cooper-Harper scale.....	4
Figure 2. Human Factors Modified Cooper-Harper for Unmanned Vehicle Display (HF MCH-UVD) evaluation scale (Cummings, et al., 2006).....	6
Figure 3. General MCH-UVD Diagnosis Tool.	8
Figure 4. UAV mission taxonomy (Nehme, et al., 2007).	15
Figure 5. UGV mission taxonomy.....	16
Figure 6. Safe mission task completion decision blocks for the General and Specific MCH-UVDs.	17
Figure 7. Flawed information retrieval descriptions for the General and Specific MCH-UVDs.....	18
Figure 8. “Mode confusion” descriptions for the General and Specific MCH-UVDs.....	18
Figure 9. Aiding mission task completion decision diamond for the General and Specific MCH-UVDs. ...	18
Figure 10. Poor attention direction descriptions for the General and Specific MCH-UVDs.....	19
Figure 11. Poor task switching support for the General and Specific MCH-UVDs.	19
Figure 12. No information aggregation descriptions for the General and Specific MCH-UVDs.....	20
Figure 13. Inefficient action implementation descriptions for the General and Specific MCH-UVDs.....	20
Figure 14. Promoting effective mission re-planning question blocks for the General and Specific MCH-UVDs.....	20
Figure 15. Insufficient decision aiding descriptions for the General and Specific MCH-UVDs.....	21
Figure 16. Inflexible decision support descriptions for the General and Specific MCH-UVDs.	21
Figure 17. Specific MCH-UVD diagnosis tool for a UGV search mission.....	22
Figure 18. Experimental Test Room.....	24
Figure 19. OPS-USERS Human Interface Workstation.....	25
Figure 20. USARSim Human Interface Workstation.	25
Figure 21. OPS-USERS Human Interface with Instruction on Creating a Search Task.....	27
Figure 22. UAV Fuel Gauge.....	27
Figure 23. USARSim Human Interface.....	28
Figure 24. Number of UAV display deficiencies identified.....	32
Figure 25. Number UGV display deficiencies identified.	33

LIST OF TABLES

Table 1: Examples of display deficiencies for different General MCH-UVD diagnosis ratings.....	13
Table 2. Experimental design.....	26
Table 3. Number of observations in general MCH-UVD categories by experimental conditions.....	30
Table 4. Number of observations in general MCH-UVD categories by level of misuse.	31

ABSTRACT

In unmanned vehicle (UV) operations, displays are often the only information link between operators and vehicles. It is essential these displays present information clearly and efficiently so that operators can interact with the UVs to achieve mission objectives. While there are a variety of metrics to evaluate displays, there is no current standardized methodology for operators to subjectively assess a display's support and identify specific deficiencies. Such a methodology could improve current displays and ensure that displays under development support operator processes. This report presents a quasi-subjective display evaluation tool called the Modified Cooper-Harper for Unmanned Vehicle Displays (MCH-UVD) diagnosis tool. This tool, adapted from the Cooper-Harper aircraft handling scale, allows operators to assess a display, translating their judgments on potential display shortcomings into a number corresponding to a particular deficiency in operator support. The General MCH-UVD can be used to diagnose deficiencies for any UV display, while the Specific MCH-UVD is UV and mission specific in its evaluation of displays. This report presents the General MCH-UVD and provides guidance on how to adapt it to create a Specific MCH-UVD through the use of UV mission taxonomies and a questioning method. A UGV search mission case study provides a how-to guide example for generating a Specific MCH-UVD. The report also presents an experiment conducted to validate the MCH-UVD and assess if a mission-specific version is necessary, or if the general form of the MCH-UVD is sufficient for different UV display evaluation. The report concludes with discussion on how to administer the scale, when a Specific scale is necessary, MCH-UVD diagnosis tool limitations, and future work.

INTRODUCTION

Current unmanned vehicle (UV) systems require interaction and supervision from human operators, who use interfaces to monitor and command UVs. Because most UV operations are conducted with operators physically separated from the UVs, interfaces often provide the only link between the operators and the UVs. It is imperative that interface displays facilitate effective and efficient information acquisition, information analysis, decision-making, and action-taking (Parasuraman, Sheridan, & Wickens, 2000). Given the rapid development of UV systems, a relatively simple but diagnostic evaluation methodology is needed to determine if these processes are supported by both proposed and current UV interfaces. Such a methodology should diagnose potential interface issues, address specific deficiencies and suggest potential remedies to improve the interfaces. An evaluation methodology assessing a display's ability to support operator processes will be valuable to evaluate and compare displays during development and testing phases to determine potential problems early. Furthermore, the use of a standardized evaluation methodology during display design and testing will aid in delivery of quality operational UV displays that assist operators' cognitive requirements in UV supervision tasks.

Standardization of a UV display evaluation methodology would ensure reliable and consistent evaluations of displays across different organizations by providing a common vocabulary for developers to characterize the effectiveness of a UV display's support of operator processes. Moreover, a standardized and validated evaluation methodology could become an inexpensive and commonly-used metric, similar to the NASA Task Load Index subjective workload rating survey (Hart & Staveland, 1988). Use of a standardized methodology could also reduce potential evaluation biases that can occur from metrics developed by the same organization that uses them.

Although a general methodology could provide a universal way to assess UV displays and is an attractive option due to simplicity in both administration and analysis, the increasing diversity and specificity of UV missions may require specific, yet standardized methodologies for interface assessment. Thus one question that needs to be addressed in the development of a standardized UV display assessment methodology is whether a more general tool like the NASA-TLX is appropriate, or if a more tailored assessment tool to a specific mission or vehicle provides more useful and accurate results. A specific methodology for display assessment would address particular aspects inherent to the vehicle and mission, which could be missed through the use of the general methodology. However, such an approach would be more costly to develop in terms of time and resources required.

This technical report presents the development of a standardized methodology for assessing UV interfaces and diagnosing their deficiencies, called the Modified Cooper-Harper for Unmanned Vehicle Displays (MCH-UVD). One particular domain, manned aircraft handling (in terms of flight control), uses a standardized methodology, called the Cooper-Harper Scale (Harper & Cooper, 1986), to evaluate aircraft handling characteristics as subjectively assessed by pilots. MCH-UVD was modeled after the Cooper-

Harper Scale by adopting its flowchart format and quasi-subjective scale. The goal behind the MCH-UVD development was to create a quasi-subjective and standardized evaluation methodology to determine and diagnose specific areas where UV displays do not support operator cognitive processes, based upon operator subjective judgments.

A General MCH-UVD scale was designed for evaluation of any UV display. The output provided from the MCH-UVD is a rating from 1 to 10. Each rating correlates to a specific UVD deficiency defined on the MCH-UVD and indicates if the display is sufficient or needs improvements. The scale is one of severity, with 1 being the perfect display and 10 being the worst display. In addition to the development of the General scale, specific scales were developed for two representative case studies in order to assess the costs and benefits of general versus specific scales. This report includes results from an experiment conducted to validate the MCH-UVD, as well as to determine if creating a mission specific MCH-UVD is necessary, or if the general form of the MCH-UVD is sufficient for different UVD evaluation.

BACKGROUND

This section presents and discusses the original Cooper-Harper Scale and the initial design of the MCH-UVD, called the Human Factors MCH-UVD scale.

Cooper-Harper Scale for Aircraft Handling Qualities

In 1957, NASA scientist George Cooper presented a rating scale, known as the Cooper Pilot Opinion Rating Scale, which attempted to quantify how pilot workload affected task performance (Cooper, 1957). Later, Robert Harper joined Cooper in modifying the original scale to address the handling characteristics of an aircraft, thus creating the Cooper-Harper Handling Qualities Rating Scheme (Harper & Cooper, 1986). Figure 1 shows the Cooper-Harper scale.

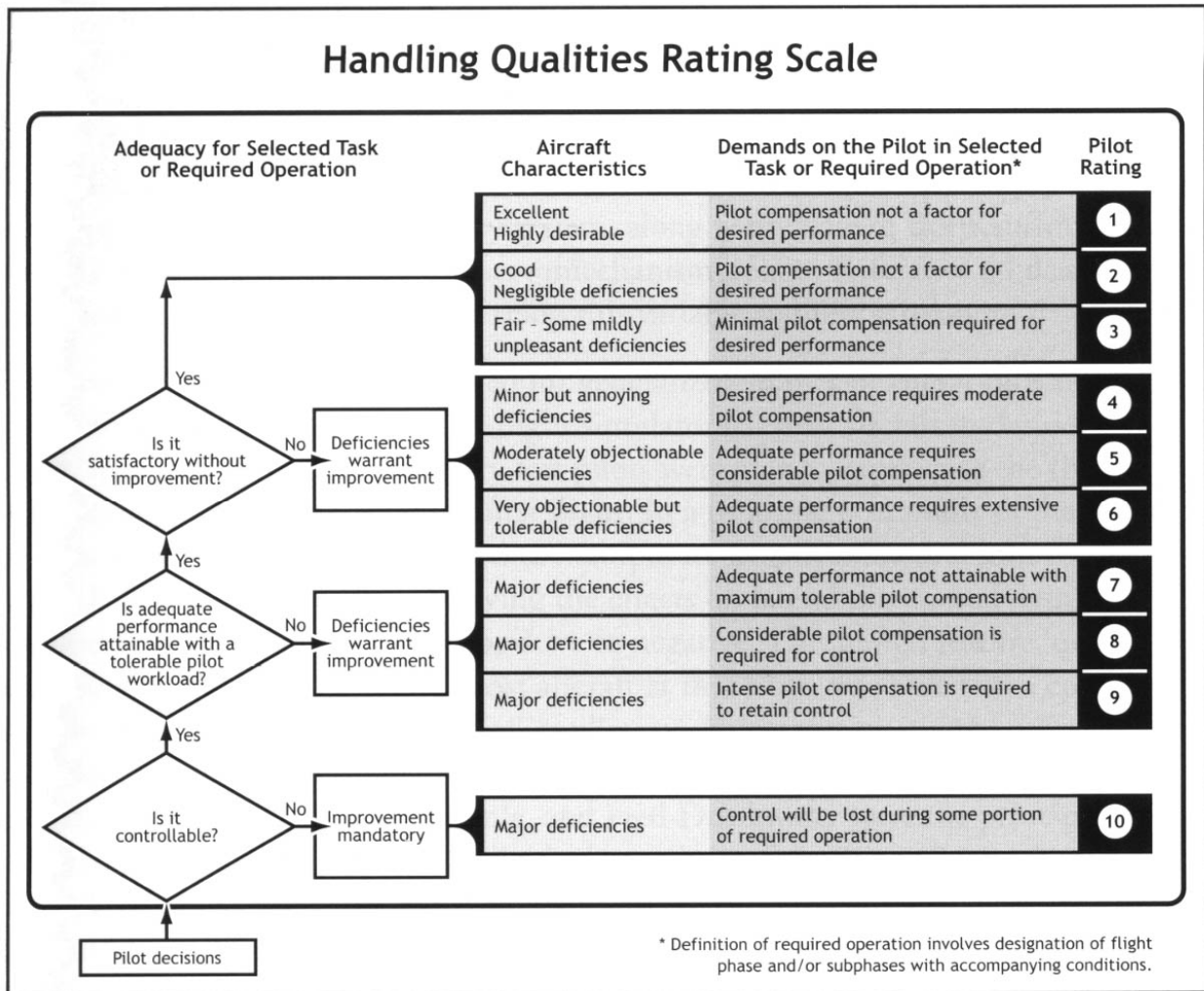


Figure 1. Original Cooper-Harper scale.

The Cooper-Harper scale allows pilots to evaluate aircraft handling qualities based upon controllability, workload, and ability to attain adequate performance. Pilots follow the flowchart and choose which descriptions of aircraft characteristics and demands best represent their handling experience. Selecting the appropriate descriptions eventually leads to a quantified number on a scale of 1 to 10, which represents the pilot's opinion of how controllable the aircraft was and how it affected his or her task performance. Because descriptive subjective pilot assessments are linked to a quantitative number, the Cooper-Harper scale is quasi-subjective. The scale's quantitative numbers, which are associated with specific descriptions of handling characteristics, can be statistically analyzed to determine the average handling aptitude of aircraft. Thus, the original Cooper-Harper scale serves as a valuable tool for aircraft handling assessment that is still used to subjectively measure aircraft design and performance (Borchers, Franklin, & Fletcher, 1998).

Modified Cooper-Harper for Unmanned Vehicle Displays

The Cooper-Harper scale for aircraft handling qualities was used as a model for development of a modified scale for UV display assessment. Unmanned vehicles differ from aircraft significantly in terms of operator control. While pilots are directly controlling and physically located in the aircraft, UV operators are often performing high-level mission management tasks and are only linked to the vehicle through displays. Thus, the original Cooper-Harper scale needed modification in order to evaluate how UV displays assist operators with higher-level mission objective tasking.

Cummings, Myers, and Scott (2006) adapted the Cooper-Harper scale to create the original Modified Cooper-Harper for Unmanned Vehicle Displays (MCH-UVD) scale. The MCH-UVD scale addressed the information acquisition, information analysis, and decision-making elements of the four-stage human information processing model (Parasuraman, et al., 2000). This model describes the cognitive steps UV operators go through when performing UV mission supervision actions. The MCH-UVD scale was designed to be presented to UV operators after display use in order to quickly assess the interface's ability to support the operator in supervising the UV and its mission. Because this initial MCH-UVD scale was generally phrased in human factors (HF) technical terminology, it is referred to throughout the rest of this report as HF MCH-UVD. Figure 2 shows the HF MCH-UVD, which served as a guide for the development of later MCH-UVD scales.

Similar to the Cooper-Harper scale, the HF MCH-UVD scale was arranged as a flowchart so it could easily be presented to operators after they used a UV display. The HF MCH-UVD ratings were separated into four distinct blocks, three of which addressed a different stage of the human information processing model (Parasuraman, et al., 2000), and one which represented acceptable display designs. Mandatory redesign was suggested for deficiencies in information acquisition or perception. Displays receiving a rating of 10 or 9, respectively, were unable to provide operators with the information that they needed, or they interfered with operator attention.

Display Qualities Rating Scale

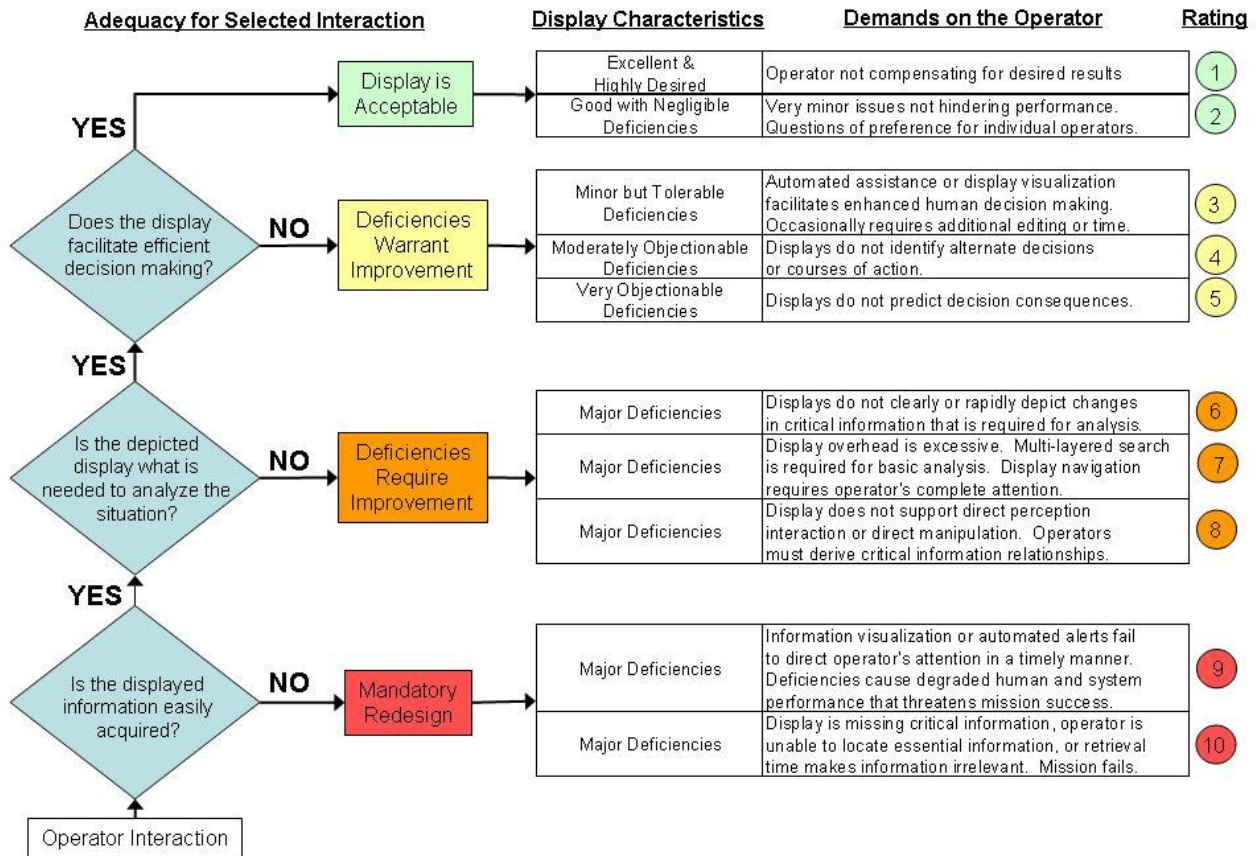


Figure 2. Human Factors Modified Cooper-Harper for Unmanned Vehicle Display (HF MCH-UVD) evaluation scale (Cummings, et al., 2006).

Displays with information analysis issues received a rating between 8 and 6. Ratings of 8 indicated display issues with perception of required information, requiring operators to determine what information sources were needed for analytical tasks and demanding mental integration of information in analysis. A display rating of 7 represented a major deficiency in cognitive resource allocation, with excessive information overhead requiring multi-layered search and navigation for analysis. The 6 rating indicated likely change blindness issues by the display, including not clearly or promptly depicting changes in critical information.

Deficiencies in decision-making were considered as warranting improvement and were addressed through ratings between 5 and 3. A rating of 5 corresponded to a display deficiency in predicting decision consequences, and a lack of addressing uncertainty. Displays receiving a rating of 4 had moderate deficiencies in representing decision-making risk by not providing alternate decisions

options. Displays providing a limited or unreliable level of automated assistance or display visualization for decision-making were considered as having minor but tolerable deficiencies and received a rating of 3.

A display was considered acceptable in information acquisition, information analysis, and decision-making if it received a HF MCH-UVD rating of 2 or 1. Ratings of 2 represented good displays with negligible deficiencies which often pertained to minor issues of preference, such as font types or sizes, or display colors. The best HF MCH-UVD rating of 1 corresponded to excellent and highly desired displays in which operators were not compensating for any display deficiencies.

Although the HF MCH-UVD scale represented a significant contribution towards evaluating UV displays, it had some deficiencies requiring redesign and refinement. First, because the HF MCH-UVD scale used Human Factors Engineering technical terminology, UV operators not familiar with the human factors field sometimes had difficulty in understanding exactly what operator demand descriptions meant, which affected their ratings. Also, the HF MCH-UVD scale did not address the action stage of the human information processing model (Parasuraman, et al., 2000), thus neglecting the effectiveness of display affordances in supporting operator tasks. Further, by prioritizing the scale strictly along the information processing model, some validity of the display deficiency severity was lost.

The HF MCH-UVD scale had been designed to rate displays across the stages of the human information processing model. However, the stages which were addressed by the scale (information acquisition, information analysis, and decision-making) were not completely divided between their representative three blocks, thus causing some overlap in the scale. For example, the HF MCH-UVD description corresponding to a rating of 7 stated that displays achieving this rating required multi-layered search for basic analysis. Searching for information applies to information acquisition, yet this deficiency was identified as relating to information analysis. Additionally, critical deficiencies threatening mission success could occur in any of the processing stages, but the HF MCH-UVD scale only suggested that mission success could be threatened with poor information acquisition.

Because of shortcomings associated with the original HF MCH-UVD scale, it was revised to create the General MCH-UVD scale, presented in the next section. This redesign included the use of more common vernacular for descriptions, with new diamond blocks more closely resembling the original Cooper-Harper scale. With this change, the human information processing model is now represented within each of the outputs from the individual diamond blocks. This change allowed the scale to address the impact of the display on mission completion, while also highlighting the actual deficiency that needs to be addressed in a redesign. The General MCH-UVD serves as a diagnostic tool for uncovering in what stage of information processing model (Parasuraman, et al., 2000) the display does not fully support operators, and classifies the severity of the deficiency in preventing mission completion.

GENERAL MCH-UVD DIAGNOSIS TOOL

The General MCH-UVD diagnosis tool, shown in Figure 3, represents a major redesign of the original HF MCH-UVD rating scale. The intent of the redesign was to represent a severity scale that defines the ability to complete the mission. Like the original Cooper-Harper scale that rated aircraft controllability on a scale of severity, the intent was to scale severity that reflected the UVD's ability to support safe mission completion. At the same time, the intent was to maintain the concepts of the human information processing model within this new scale, as this is a critical component to UV display designs.

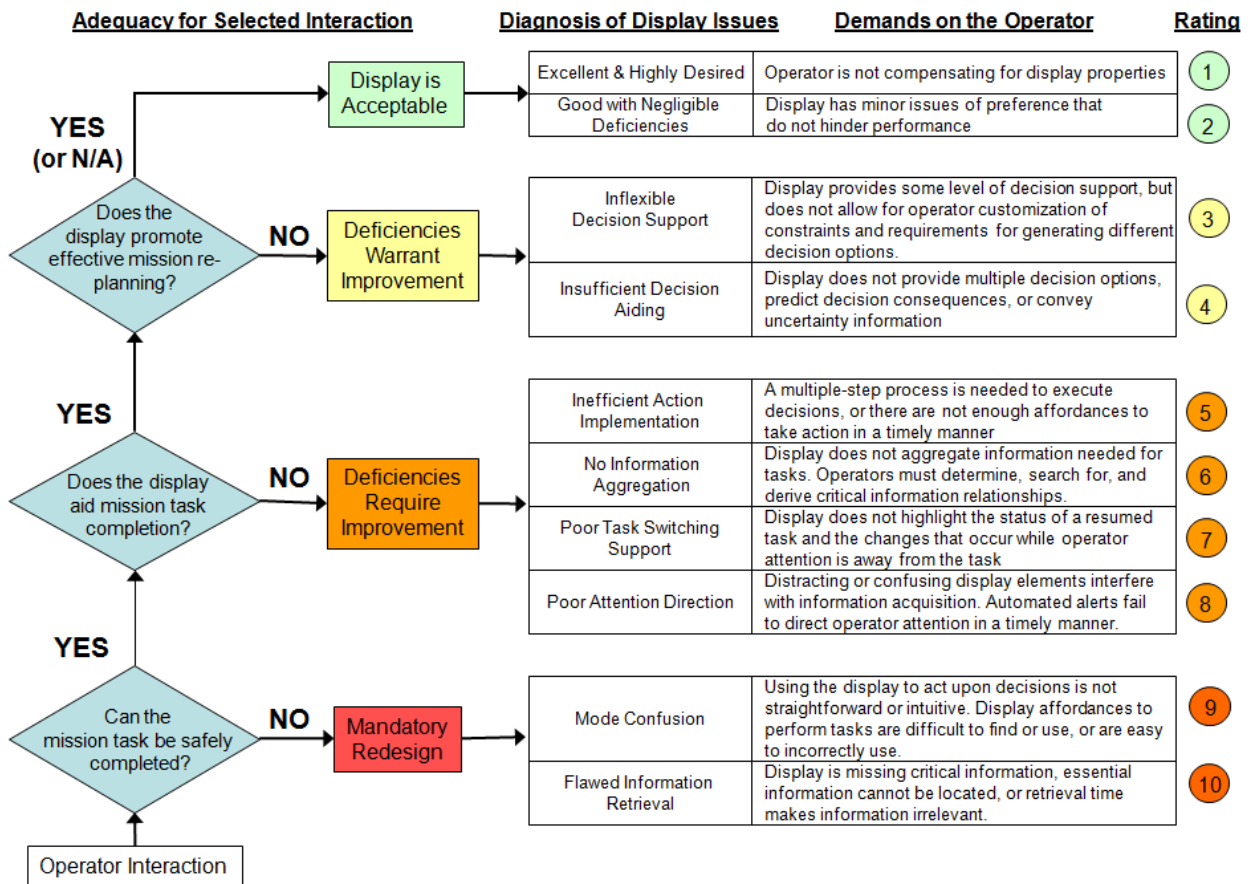


Figure 3. General MCH-UVD Diagnosis Tool.

Similar to the HF MCH-UVD, the General MCH-UVD tool uses a flowchart format, which guides UV operators to determine display deficiencies. The diamond block questions on the left of the tool ask operators if the UV display supports completing the mission safely, if it aids in mission completion, and for applicable cases whether it aids mission re-planning. Based upon an operator's answer (yes or no, respectively), the tool guides them to another question querying about a higher stage of supervisory support or directs them to descriptions of potential information processing problems specific to that level.

Within the individual diamond groups in Figure 3, the human information processing model is applied on a severity scale, from information acquisition to information analysis to decision-making and finally action-taking tasks. The deficiencies are deemed to be more severe for tasks earlier in the human information processing model, because if the display is flawed in supporting an earlier stage, it does not matter how good the display supports the later phase because the operator is acting upon a flawed input.

When operators are directed to the right of the diamond block questions, they examine a set of descriptions pertaining to potential issues and the relative demands those issues have on the operator. Within each diamond block, operators choose between two to four different descriptions, each of which corresponds to a number. Thus, the General MCH-UVD diagnosis tool is quasi-subjective, allowing for statistical analysis of operator opinions on UV displays.

Current UV displays are predominantly visual displays with embedded auditory alerts (Dixon, Wickens, & Chang, 2005). As UV operations shift from multiple operators controlling one UV to one operator controlling multiple UVs (Barbato, Feitshans, Williams, & Hughes, 2003; Culbertson, 2006; Tsach, Peled, Penn, Keshales, & Guedj, 2007), multi-modal displays will likely become more prevalent to reduce operator workload, by offloading some cognitive processing to other senses such as auditory and haptic. Thus, MCH-UVD diagnosis tool is not specific to visual only displays, but is designed to aid operators to evaluate displays in different modalities including visual, auditory, haptic, and multimodal displays.

The next sections discuss in detail the general display problems a UV display could experience, which correspond to each General MCH-UVD diagnosis tool rating.

Decision Diamond #1: Cannot Complete Mission Task Safely

When the mission task cannot be safely completed, display redesign is mandatory. Two UV display deficiencies flow from a failure to safely complete the mission task: flawed information retrieval and mode confusion.

Flawed Information Retrieval – Rating 10: A display is considered to be flawed in information retrieval when it is missing critical information, essential information cannot be located, or information becomes outdated because of long retrieval times. A display that requires extensive multi-layered search could receive a rating of 10 if searching creates long retrieval times. Generally, under this diagnosis displays do not provide operators with the necessary information they need for tasks, making higher-level information processing virtually impossible and increasing the likelihood of UV mission failure.

For example, one of the major causes of the Three Mile Island disaster was the lack of display indication of a relief valve failure (Rubinstein & Mason, 1979). The displays indicated a rise in the level of coolant rather than presenting the underlying system malfunction. Based on this flawed information, the

operators concluded that the emergency coolant flow should be reduced, which resulted in two critical hours wasted implementing an incorrect course of action that resulted in catastrophic consequences.

Mode Confusion – Rating 9: UV display which do not provide straightforward or intuitive ways to act upon decisions are classified as having confusing action implementation. These displays, which receive a General MCH-UVD rating of 9, may have display affordances that are difficult to find or use, or that are easy to use incorrectly, thus making operator tasks hard to perform or easy to erroneously execute (Bredereke & Lankenau, 2002). Incorrect task performance because of poor affordances could threaten mission success, even when information acquisition, analysis, and decision-making have been efficiently and properly performed.

For example, in 2006, an MQ-9 Predator B impacted the terrain northwest of Nogales, Arizona during a nighttime border patrol mission and destroyed the aircraft. A Human Factors Analysis and Classification System examination conducted by Carrigan, Long, Cummings, & Duffner (2008) identified mode confusion to be one of the major factors leading to this crash. For this flight, there were two nearly physically identical consoles for the pilot and the payload operator, with the ability to transfer control from one console to the other. However, the functionality differed vastly depending on the mode of operation. The throttle quadrant on the console provided UAV throttle function for piloting the aircraft, and operated camera tracking settings for payload operation. Before transferring control from one console to the other, the condition lever position on the payload operator console had to be matched to the current positioning on the pilot console. Initially, the pilot console lever was set to allow fuel flow (forward position) and the payload operator console lever was set to lock the iris (center position). After the transfer, the iris setting became the new engine setting, causing the engine to shut down. The mishap pilot was unaware of the mismatched control situation which led to mode confusion, and ultimately, the loss of the UAV.

Decision Diamond #2: Display Does Not Aid Mission Task Completion

The display may allow the user to complete the mission safely, but it still may not help the operator complete the mission task. There are four deficiencies that can result from failure to pass this diamond block: poor attention direction, poor task switching support, no information aggregation, and inefficient action implementation. These deficiencies require improvement, because though the operators may be able to safely complete the mission, the displays are not supporting them. With display redesign in these cases, the operators may be able to complete their mission task in a more efficient way.

Poor Attention Direction – Rating 8: A display receiving a rating of 8 provides operators with the information they need for tasks, but contains distracting display elements or clutter which sometimes interferes with efficient information acquisition. Additionally, if the display uses automated alerting, these alerts do not attract operator attention in a timely manner. It should be noted that this rating does not state

that all displays should use automated alerts, but if a display does, they should effectively attract operator attention. Under a rating of 8, operator information acquisition is hindered, but is still possible.

An illustrative case is the cruise ship Royal Majesty accident in 1995. After about 35 hours in transit from Bermuda to Boston, the ship grounded on the Nantucket shoals (NTSB, 1995). Shortly after departure, GPS switched to dead-reckoning mode as it was no longer receiving satellite signals. None of the crew members were aware that GPS was in dead reckoning mode for the duration of the trip. There was a very small visual indication on the GPS displaying dead reckoning, cluttered by all the other visual information. Moreover, the aural alarm indicating the switch to dead reckoning sounded only for a brief time in the chart room, not on the bridge central console where the watch officer stood. Thus, neither the visual display, nor the auditory alert directed crew attention to the change in the system state.

Poor Task Switching Support – Rating 7: When the operator intentionally moves attention to one task to another, or when he or she is interrupted, switching to the new task and switching back to the previous task comes with a cost (McFarlane, 2002; Rogers & Monsell, 1995; Sheridan, 1972). UV displays receiving a rating of 7 have issues with supporting task switching. Although these displays in part support information analysis tasks, they do not clearly highlight the status of the resumed task and the changes that have occurred while the operator's attention was away from the task. Display issues with task switching support can cause operators to make decisions and take actions based upon incorrect information, and may increase the time operators spend on analysis tasks. A classic human factors example of this problem is checklist interruption. There have been a number of aircraft accidents that have occurred due to such interruptions, particularly in the takeoff configuration when pilots are interrupted in the takeoff checklist process and forget to lower the flaps, resulting in stall and crash of the aircraft.

No Information Aggregation – Rating 6: UV displays that do not amass task information collectively or require operators to determine, search for, and derive critical information relationships are considered deficient in information aggregation. Displays receiving a rating of 6 do not suggest what information to analyze for tasks and do not co-locate different pieces of information related to a specific task. This increases the cognitive processing of operators who need to determine what information to analyze, where the information is, and how to analyze it. These problems are exacerbated under time-pressure and dynamic settings, which are inherent characteristics of UV domains.

Inefficient Action Implementation – Rating 5: A UV display receives a rating of 5 if it has inefficient action implementation. These displays either require an unnecessary multiple step process to execute actions, or do not provide enough affordances to take action. In this case, displays generally support operator actions, although perhaps not efficiently, which could have negative effects on a UV mission, particularly under time-critical situations.

Decision Diamond #3: Display Does Not Promote Effective Mission Re-planning

Mission re-planning technology and support is not necessary in all UV missions, which is why there is a not applicable (N/A) option for this diamond. This allows the evaluator to bypass the question in cases where this functionality is not required. If it is, there are two deficiencies a UVD may have: insufficient decision aiding and inflexible decision support. These deficiencies only warrant improvement because they are not as critical in supporting the human operator as the deficiencies presented before. They are generally a supplemental tool, which could enhance the operator's interaction with the display.

Insufficient Decision Aiding – Rating 4: UV displays which operators do not consider to aid them in decision-making receive a rating of 4. These displays do not provide multiple decision options or predict the potential consequences of decisions. Additionally, UV displays receiving a rating of 4 do not convey uncertainty information about decision alternatives, their potential consequences or about decision-making in general. Insufficient decision aiding by a UV display can make it difficult for operators to make and understand the effects of decisions, increasing the likelihood of operator error and potentially jeopardizing mission success.

Inflexible Decision Support – Rating 3: Operators who rate a UV display as a 3 believe that it provides some level of support in decision-making, but the display does not allow for operator customization of decision-making constraints and requirements to narrow down decision options. This inflexible decision support, as subjectively deemed by operators, is useful to the decision-making process but may not help operators make optimal decisions, which could have potentially negative effects on a UV mission, particularly in time-critical situations.

Acceptable Displays

Displays which support completing the mission safely, aid in mission completion, and for applicable cases, aid mission re-planning are considered acceptable displays for UV mission use. These displays, however, may have a few, but negligible, deficiencies that do not hinder operator performance in supervising the UV mission. Thus, acceptable displays can be classified as good with negligible deficiencies, or excellent and highly desired.

Good Displays with Negligible Deficiencies – Rating 2: UV Displays which receive a General MCH-UVD rating of 2 support human information processing through all four stages, but have very minor issues of preference that do not hinder operator performance. Example issues include preference of font style or size, display colors, or display element arrangement or sizing. These minor deficiencies should not affect an operator's ability to acquire and analyze information, make decisions, or take action to supervise the UV mission.

Excellent and Highly Desired Displays – Rating 1: A UV display receives a rating of 1 when the operator can perform all four stages of information processing and is not compensating for any deficient display properties. This rating signifies that the display fully supports information acquisition, analysis, decision-making, and action-taking, assisting the operator in optimally supervising the UV mission.

Summary of the General MCH-UVD Diagnosis tool

For each of the 10 General MCH-UVD diagnosis tool ratings, Table 1 gives examples of what display attributes could cause an operator to choose a particular rating for a display. The General MCH-UVD diagnosis tool improves upon the HF MCH-UVD display qualities rating scale in that it uses more common verbiage and makes the MCH-UVD more like the original Cooper-Harper scale so that rating a display on this scale not only identifies the design deficiencies, but classifies the severity of the problem.

Because of its universal format, the General MCH-UVD tool could be used to evaluate and diagnose UV displays pertaining to any type of vehicle or mission. However, there could be instances when display developers may want to evaluate displays specific to a certain UV and mission type. By using a MCH-UVD diagnosis tool which is UV and mission specific, display developers would be able to pinpoint not only the severity of display deficiencies, but also what UV and mission-specific elements the display is not supporting. The next section of this report presents a methodology for generating mission specific MCH-UVDs.

Table 1: Examples of display deficiencies for different General MCH-UVD diagnosis ratings.

Rating	Examples of Display Deficiencies
1	The operator has no issues of preference and is able to complete the mission safely with efficient support from all displays.
2	The operator prefers different orientation of display elements.
3	Operator cannot enter time constraints when making schedule management decisions for a UV mission.
4	The display does not convey the likelihood of success for a decision.
5	A four step process is required to alter a UV path, which requires access to several sub-menus.
6	UV location and velocity information are on separate screens.
7	Significant changes in UV health are not salient on the display.
8	The operator is not alerted when a UAV deviates off-course.
9	The affordance for firing a time-critical missile makes it appear that the missile has fired, when in fact it is only been armed and requires an additional step to fire it.
10	Display does not provide any velocity information for a UAV.

GENERATING SPECIFIC MCH-UVDs

This section describes the method through which UV and mission-specific MCH-UVD diagnosis tools (henceforth simply referred to as Specific MCH-UVDs) can be created. In order to develop a method for designing UV and mission-specific MCH-UVD diagnosis tools, missions for two types of UVs, UAVs and UGVs were identified and taxonomies were developed. Defining the mission is a critical first step for a mission-specific MCH-UVD, as a clear mission definition allows developers to populate the specific MCH-UVD scale with detailed information in order to better evaluate displays supporting that mission. It should be noted that both these UAV and UGV mission taxonomies could change with the development of new technologies and UV capabilities.

Unmanned Vehicle Mission Taxonomies

Nehme, Crandall, & Cummings (2007) developed a taxonomy for UAV missions, shown in Figure 4. This taxonomy consists of a tree-like structure and has three levels. The main UAV missions are intelligence/reconnaissance, drones, transport, extraction, insertion, communication (voice and data), and surveillance. The second and third levels of the taxonomy give more specific details of the missions, thus further classifying them. Specific MCH-UVD scale developers creating a scale for a UAV mission can use this taxonomy to determine exactly what the UAV's mission is and any important details pertaining to it. More information on the UAV mission taxonomy, including mission type definitions, can be found in Nehme et al. (2007).

In order to support development of UGV mission-specific MCH-UVD diagnostic tools, a similar UGV mission taxonomy was created (Figure 5). As in the UAV mission taxonomy, the UGV taxonomy has a tree-like structure and consists of three levels (the middle section of Figure 5 shows the second level, with some branches offset due to space constraints). The main UGV missions are logistics, payload delivery, decoy, intelligence/surveillance/reconnaissance/sensing, search and rescue, communication, and neutralization. Similar to the UAV mission taxonomy, the second and third levels further classify and describe the first level UGV missions. Appendix A contains the UGV mission taxonomy definitions along with examples of current UGVs performing certain mission types (GlobalSecurity.org).

The taxonomies presented in Figures 4 and 5 present one way to identify a mission for a particular UV, which is the critical first step in developing criteria to assess a display if specificity is of concern. The important result of this step is the identification of the mission, at the level of granularity needed to generate specific display concerns. Developers can use these UAV and UGV mission taxonomies to help specify mission elements, and they can also draw upon mission requirements, objectives, and task lists.

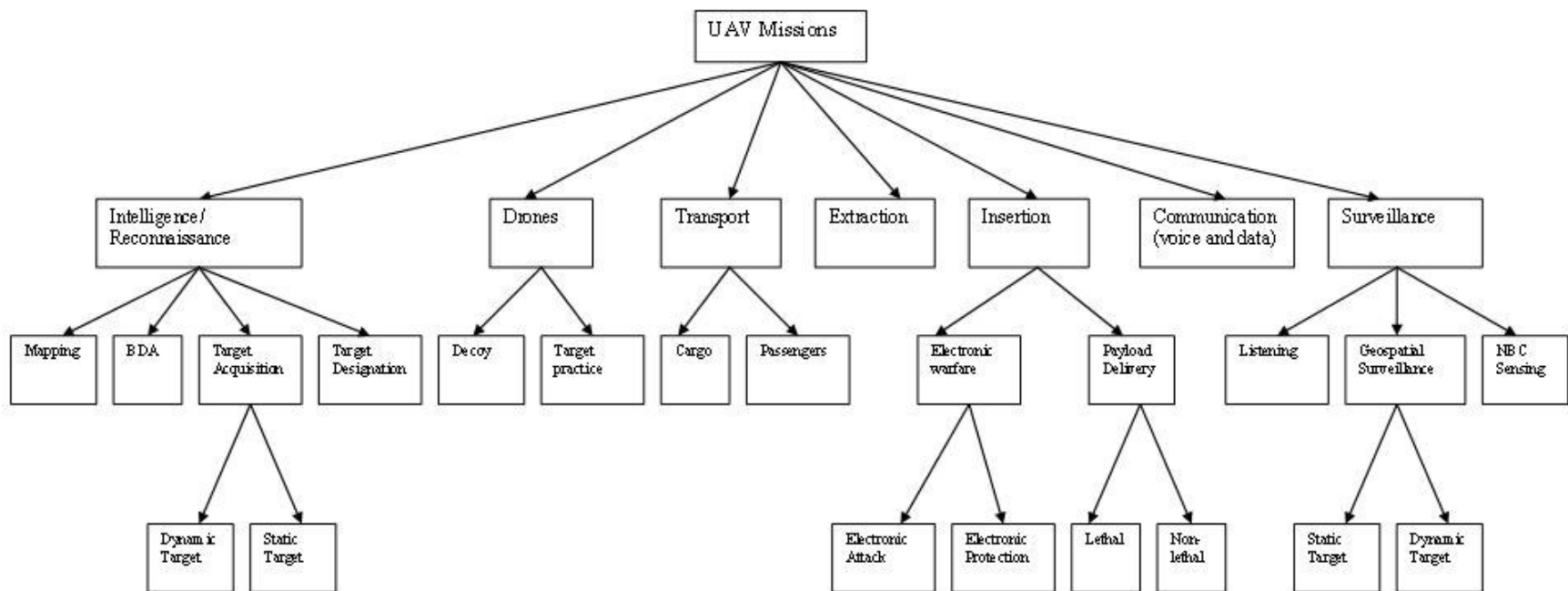


Figure 4. UAV mission taxonomy (Nehme, et al., 2007).

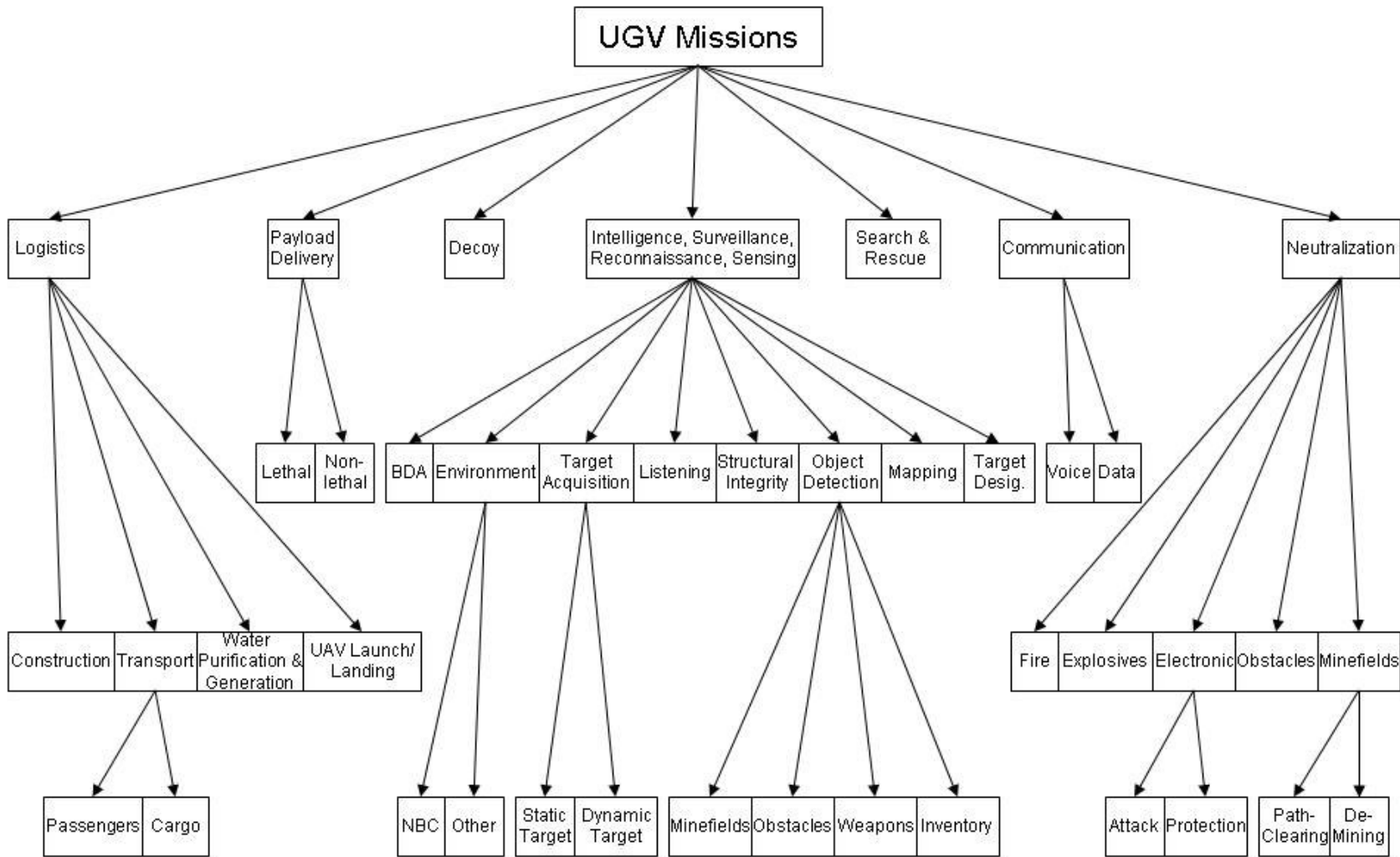


Figure 5. UGV mission taxonomy.

Guiding Principles of Specific MCH-UVD Tool Generation

The following guide gives step-by-step instructions on how to create a Specific MCH-UVD scale. Throughout the guide, a Specific MCH-UVD diagnosis tool for a UGV search mission serves as an example. In generating Specific MCH-UVD scales, the General MCH-UVD diagnosis tool is used as a template, and the figures in this section highlight the differences between the Specific and the General Scales by underlining the relevant descriptions on both of the scales. In order to create the specific scales, we propose that a set of questions should be asked about the UV mission in terms of display support. Answers to these questions are then inserted into the General MCH-UVD scale to replace its broad descriptions.

Decision Diamond #1: Cannot Complete Mission Task Safely

Safe Mission Task Completion Decision Diamond: To create the safe mission task completion task decision diamond, developers should ask themselves the following:

- What is the specific mission task that needs to be safely completed by this UV?

This decision diamond should not necessarily go into detail about specifics of the mission, but should give enough description that it is clear what a failed mission would be. Figure 6 shows the General decision block and the example UGV search Specific MCH-UVD complete mission task decision block.

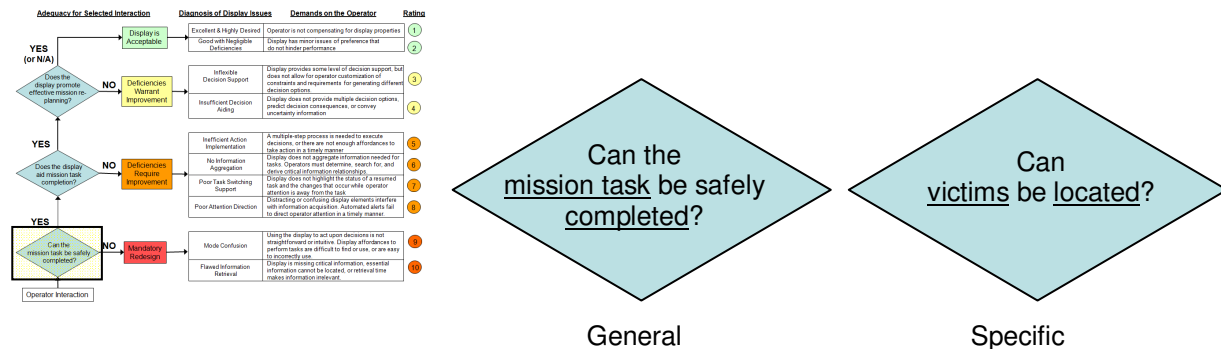


Figure 6. Safe mission task completion decision blocks for the General and Specific MCH-UVDs.

Flawed Information Retrieval – Rating 10: The question that needs to be answered for this level is:

- What is the critical information needed for the mission?

In contrast to the information acquisition question in the general case, this description should go into more depth about specifically what mission information is critical. In Figure 7, the specific mission critical information (UGV state, search environment, victims) replaces the General scale’s broad reference to “critical information.”

Display is missing critical information, essential information cannot be located, or retrieval time makes information irrelevant.

General

Critical information about UGV state, the search environment, or victims and their locations is missing, difficult to find, or long retrieval time makes information irrelevant.

Specific

Figure 7. Flawed information retrieval descriptions for the General and Specific MCH-UVDs.

Mode Confusion – Rating 9: The following questions should be answered:

- What are the specific critical mission tasks that the operator should perform?
- What specific actions do operators need to perform during the mission?

The specific actions that operators need to perform during the mission may come from objectives or tasking lists, and will likely correspond to the decisions operators make in the mission. Figure 8 presents the General and Specific UGV search MCH-UVD diagnosis tools for this rating.

Using the display to act upon decisions is not straightforward or intuitive. Display affordances to perform tasks are difficult to find or use, or are easy to incorrectly use.

General

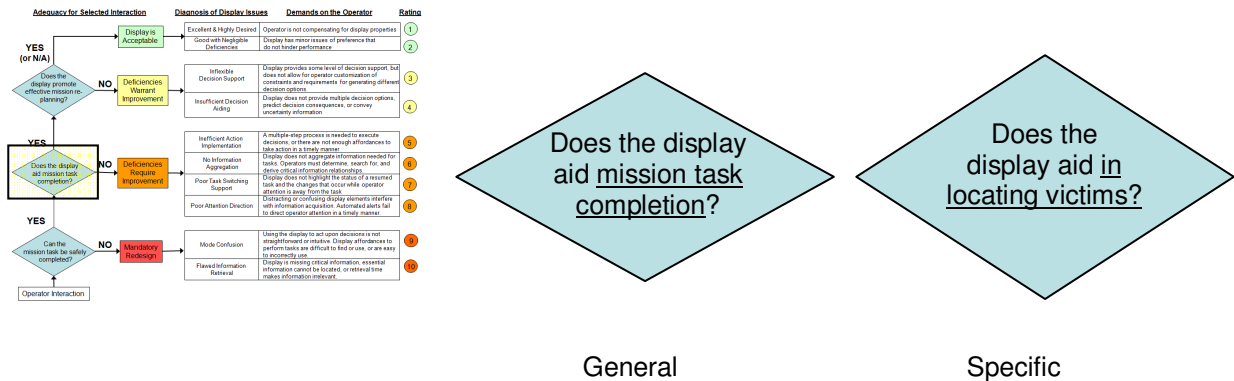
Using the display to change search patterns or to locate victims is not straightforward or intuitive. Display affordances to perform the search and to identify the locations of victims are difficult to find or use, or are easy to incorrectly use.

Specific

Figure 8. “Mode confusion” descriptions for the General and Specific MCH-UVDs.

Decision Diamond #2: Display Does Not Aid Mission Task Completion

Aiding Mission Task Completion Decision Diamond: This block should be updated with the specific mission task that was identified previously (Figure 9).



General

Specific

Figure 9. Aiding mission task completion decision diamond for the General and Specific MCH-UVDs.

Poor Attention Direction – Rating 8: To update this level, the following two questions should be answered:

- What are the entities that require operator attention?
- Are there and should there be any automated alerts?

Potential automated alerts in a UGV search mission could include obstacle detection or identification of potential victims. It should be noted that some missions may not require automated alerts, in which case this part of the description could be removed from the Specific MCH-UVD scale. However, potential inadvertent omission of critical alert states should also be noted. Figure 10 shows a side by side comparison of this block for the General and Specific MCH-UVDs.

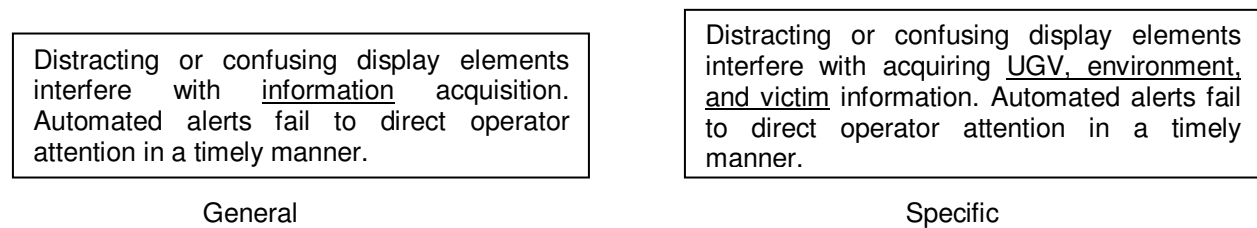


Figure 10. Poor attention direction descriptions for the General and Specific MCH-UVDs.

Poor Task Switching Support – Rating 7: This block is updated with the specific entities that require operator attention, which were previously identified (Figure 11) with the question:

- What are the entities that require operator attention?

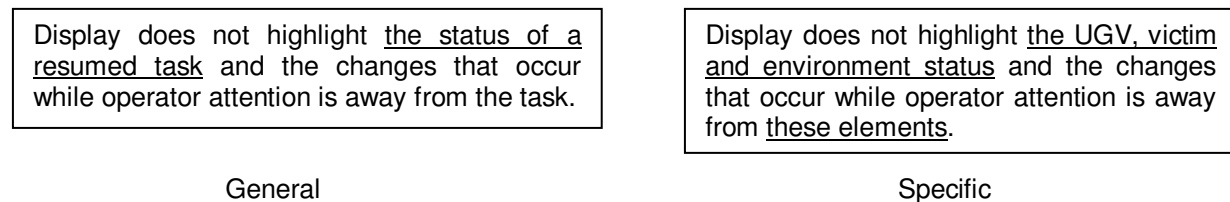


Figure 11. Poor task switching support for the General and Specific MCH-UVDs.

No Information Aggregation – Rating 6: To create the Specific MCH-UVD no information aggregation description, the following question should be answered:

- Which tasks are related, and what information is needed to indicate these relationships?

As shown in Figure 12, the related mission tasks fill in the first sentence of the Specific MCH-UVD description, while the relationship information completes the second. While this block does not directly state what the relationships are, it should articulate what information is involved in those relationships.

Display does not aggregate information needed for tasks. Operators must determine, search for, and derive critical information relationships.

General

Display does not aggregate information needed for searching for and locating victims. Operators must determine, search for, and derive critical information relationships between UGV location and the locations of victims.

Specific

Figure 12. No information aggregation descriptions for the General and Specific MCH-UVDs.

Inefficient Action Implementation – Rating 5: The specific actions the operators need to perform during the mission, which were previously identified, is used to update this block (Figure 13). Thus, the following question is used again:

- What specific actions do operators need to perform during the mission?

A multiple-step process is needed to execute decisions, or there are not enough affordances to take action in a timely manner.

General

A multiple-step process is needed to change the search patterns or to locate victims, or there are not enough affordances to take action in a timely manner.

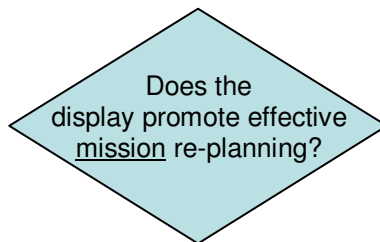
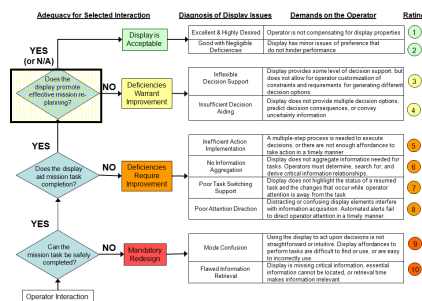
Specific

Figure 13. Inefficient action implementation descriptions for the General and Specific MCH-UVDs.

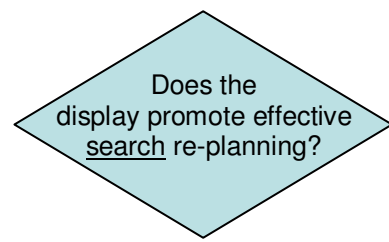
Decision Diamond #3: Does Not Promote Effective Mission Re-planning

Promoting Effective Mission Re-Planning Question Block: To create the Specific MCH-UVD promoting effective mission re-planning decision diamond, developers should ask themselves the following:

- What is the specific mission task that may require re-planning?



General



Specific

Figure 14. Promoting effective mission re-planning question blocks for the General and Specific MCH-UVDs.

The goal of this decision diamond is to allow for advanced decision support tools to be used in UVD designs. It is not necessary or applicable for all missions, which is why the diamond block has the N/A option to bypass it. Figure 14 shows the General and Specific UGV search decision-making decision diamond, highlighting where changes may exist depending on how specific one wants to be in evaluating various support tools.

Insufficient Decision Aiding – Rating 4: This block should be updated with the answers to the following questions:

- What critical decisions need to be made during mission replanning?
- What kinds of uncertainty exist in decision-making?

Figure 15 shows the General and Specific UGV search insufficient decision aiding descriptions.

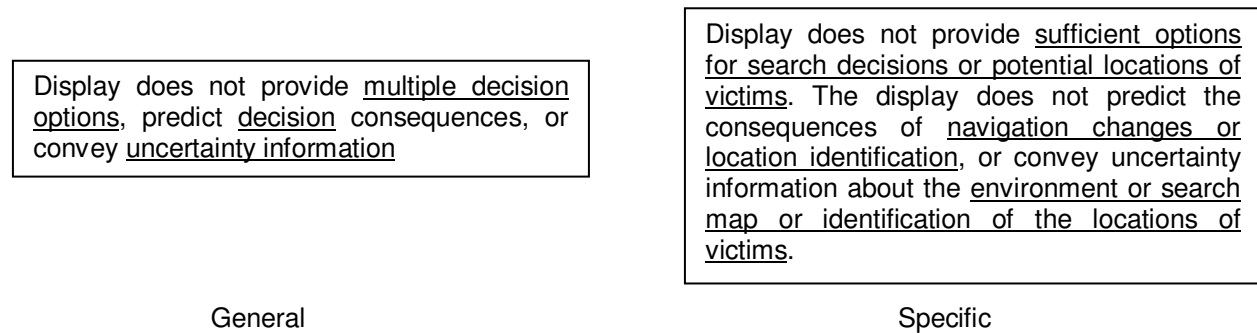


Figure 15. Insufficient decision aiding descriptions for the General and Specific MCH-UVDs.

Inflexible Decision Support – Rating 3: In designing the Specific MCH-UVD inflexible decision support description, developers should ask themselves the following about their UV and its mission:

- What kinds of constraints or requirements could help generate decision options?

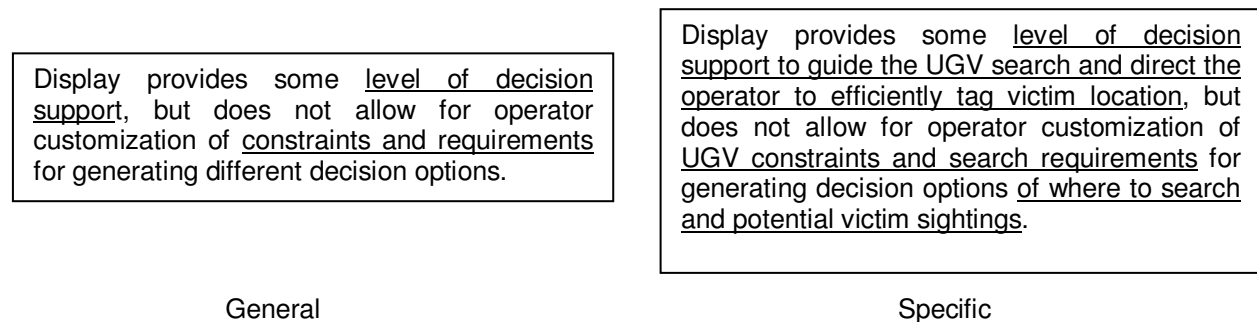


Figure 16. Inflexible decision support descriptions for the General and Specific MCH-UVDs.

This question prompts Specific MCH-UVD scale developers to consider what constraints and requirements operators may want to manipulate in order to generate decision options. These may be general or very specific depending on the level of mission specificity. In the UGV search mission

example, it may be useful for an operator to be able to specify UGV constraints and use search requirements to generate decision options (Figure 16).

Acceptable Displays

The descriptions for rating 1 (excellent and highly desired displays) and rating 2 (negligible deficiencies) remain unchanged, since neither of these ratings pertains to elements which could differ due to UV type or mission.

Example Specific MCH-UVD Scale for UGV missions

Figure 17 shows a Specific MCH-UVD diagnosis tool for the UGV search mission case study consisting of the question blocks and operator demands previously described.

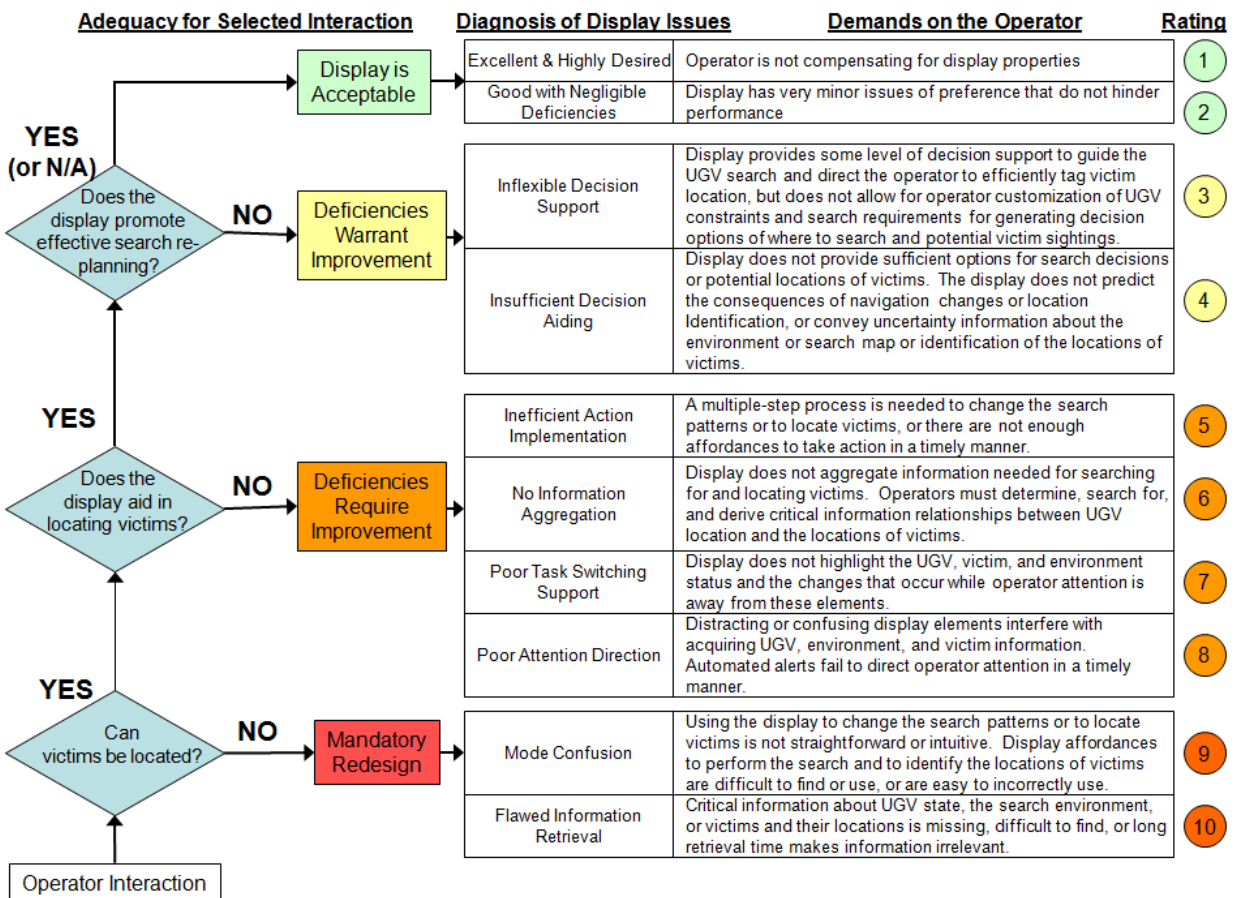


Figure 17. Specific MCH-UVD diagnosis tool for a UGV search mission.

As compared with the General MCH-UVD scale (Figure 3), the UGV search scale is more detailed with longer descriptions. However, these specific descriptions can help identify display deficiencies related to the UGV search mission, determining not only where a display poorly supports operator information processing, but also what specific mission elements are not supported. The advantage of the Specific MCH-UVD scale is that it can pinpoint display deficiencies with respect to specific mission elements, thus reducing potentially major redesigns of a display.

VALIDATION EXPERIMENT

During the development of MCH-UVD, a human subject experiment was conducted to evaluate both general and specific scales using two different UV displays. This experiment helped us better define the ratings, and also assess the construct validity of MCH-UVD by investigating the correlation between MCH-UVD ratings and display misuses. The comparison of General and Specific scales was deemed important since the potential benefits of creating a Specific Scale may not surpass the cost of doing it. The MCH-UVD was administered to participants as a post-test survey for evaluating two types of UVDs: UAV, and UGV, presented in its general or specific form.

Participants

Sixty participants completed the study. The participants consisted of 24 females and 36 males, ages ranging from 18 to 45 years (mean = 20.7, standard deviation = 4.01). The experiment took between 1 and 1.5 hours to complete. The participants were compensated monetarily at \$10 per hour and were eligible to win a gift card based on their performance with the UVD. The top performing participants in each UVD condition received a \$50 gift card to Best Buy[®].

Apparatus

The experiments were conducted in an experimental space at the Massachusetts Institute of Technology's Humans and Automation Laboratory (HAL) (Figure 18).



Figure 18. Experimental Test Room.

The UAV condition utilized the Onboard Planning System for UAVs in Support of Expeditionary Reconnaissance and Surveillance (OPS-USERS) simulator, developed by Aurora Flight Sciences. OPS-USERS simulates multiple unmanned vehicles conducting search and tracking operations in various

environments. The UGV condition was carried out on Urban Search and Rescue Simulator (USARSim), an open source, high fidelity simulation of urban search and rescue robots and environments (Lewis, Wang, & Hughes, 2007). The simulations were run on a Dell Optiplex GX520 with a Pentium 4 processor connected to a Smart Board display (seen in the left of Figure 18). The OPS-USERS simulation used an additional computer, a Dell Dimension 9100 with a Pentium 4 processor, and equipped with three 17 inch desktop displays. Through this computer, the test proctor was provided with a GUI, which showed the locations of different targets being searched by the participants.



Figure 19. OPS-USERS Human Interface Workstation.



Figure 20. USARSim Human Interface Workstation.

The UGV and UAV human interfaces were run on Dell Dimension 5150s with Pentium 4 processors, NVIDIA Quadro NVS graphics cards, and three 17 inch desktop displays. The OPS-USERS human interface was run on the upper left screen displaying in 1600 x 1024 pixels and 32 bit resolution (Figure 19). The human interface for USARSim (Wang & Lewis, 2007) was run on the lower left screen displaying in 1280 x 1024 pixel and 32 bit color resolution (Figure 20). Participants controlled the simulators through a generic corded computer mouse.

Experimental Design

The experiment was a 2x2 completely randomized design (Table 2). The independent variables were UVD type (UAV, UGV), and MCH-UVD type (general, specific). Participants were randomly assigned to one of the four conditions. The UAV condition utilized the OPS-USERS simulation, whereas the UGV condition used USARSim. As presented in the previous sections, the general MCH-UVD (Appendix B) was modified to create two specific MCH-UVDs: UAV-specific (Appendix C) and UGV-specific (Appendix D).

It should be pointed out that the scales evaluated in the experiment were earlier versions and were slightly different than the final scales presented in the previous sections. Based on the experimental findings, presented in the next chapter, rating number seven was modified from “change blindness” (i.e., inability to detect a change) to “poor task switching support”. Both the “change blindness” and the “poor task switching support” relate to attention allocation. However, “change blindness” has a large overlap with rating eight, that is, “poor attention direction”. Thus, this modification was performed to more clearly distinguish the display deficiencies represented in ratings seven and eight.

Table 2. Experimental design.

UV Display Type	MCH-UVD Type	
	General	Specific
UAV	n = 15 (group A)	n = 15 (group B)
UGV	n = 15 (group C)	n = 15 (group D)

Experimental Tasks

In this experiment, participants performed in a supervisory role, executing their respective UAV or UGV missions.

UAV mission/display: The operators in the UAV condition performed a dynamic target acquisition mission. OPS-USERS, for this experiment, allowed the operator to search for and track targets with a single UAV while monitoring the UAV’s flight path and event timeline. The UAV was designed to search for targets and then track the targets upon target identification. Participants were instructed to monitor a

water canal for passing ships. The objective was to maximize the number of ships found and the amount of time the ships were tracked.

The human interface (Figure 21) provided a map for tracking the UAV on its present flight path, along with an interaction control panel (bottom) for making UAV control inputs, and a panel for accepting or rejecting automation plans (left). Control inputs included creating a search or tracking task for the UAV, setting task start and end times, and creating task duration. For the purposes of this experiment, the operators were allowed to assign the UAV one task at a time. Thus, both the task value and the delay penalty seen in Figure 21 were always set to high. The operator could also modify and delete tasks using the control panel. The flight time and range of the UAV was limited by the fuel tank, so the operator had to be aware of the fuel level and periodically allow the UAV to return to base to refuel. The green bar above the UAV in Figure 22 is the fuel gauge. The green bar turned to red when the fuel was low.

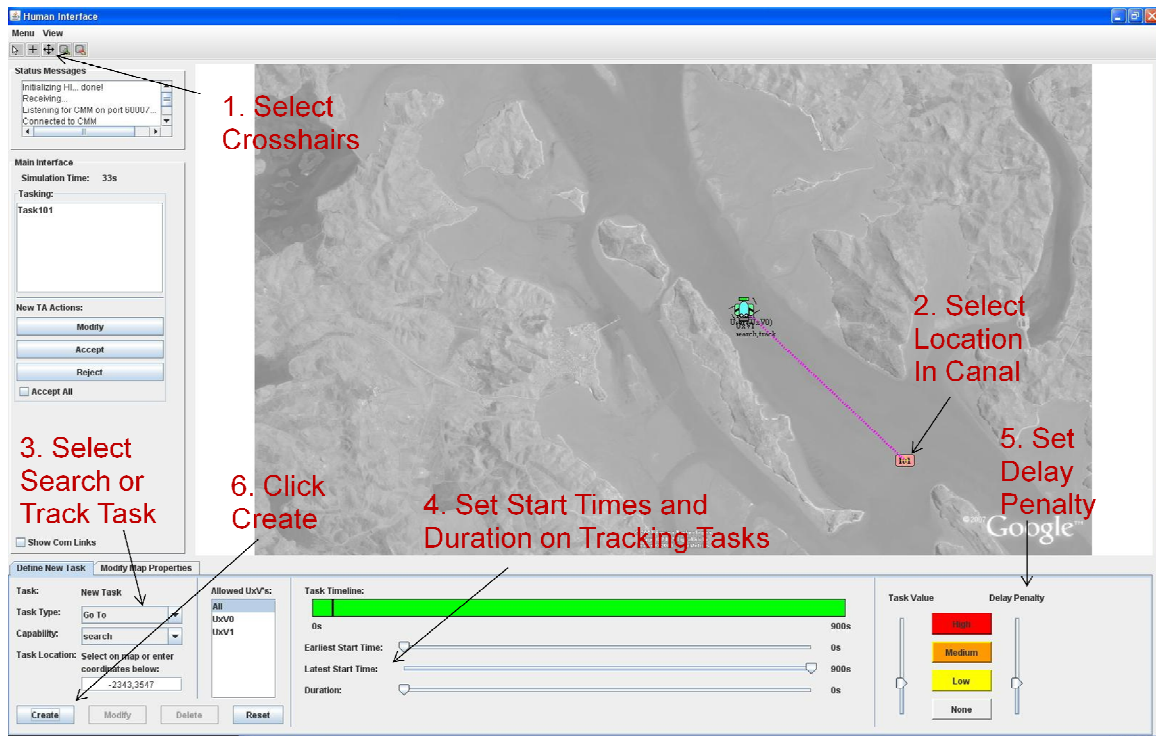


Figure 21. OPS-USERS Human Interface with Instruction on Creating a Search Task.



Figure 22. UAV Fuel Gauge.

To search the canal for a target, the operator had to select a location in the canal and create a search task, and then accept the plan that the automation presented to them (Appendix E). Once the operator accepted the plan, a thin green line showing the projected flight path from the current location of

the UAV to the task showed up and the UAV began to fly to the task. Once the UAV flew over a ship (target) in the canal, the ship appeared on the map and an initial tracking task was automatically presented to the operator (Appendix E). The operator then had to accept the initial tracking task and the UAV tracked the ship for the specified duration.

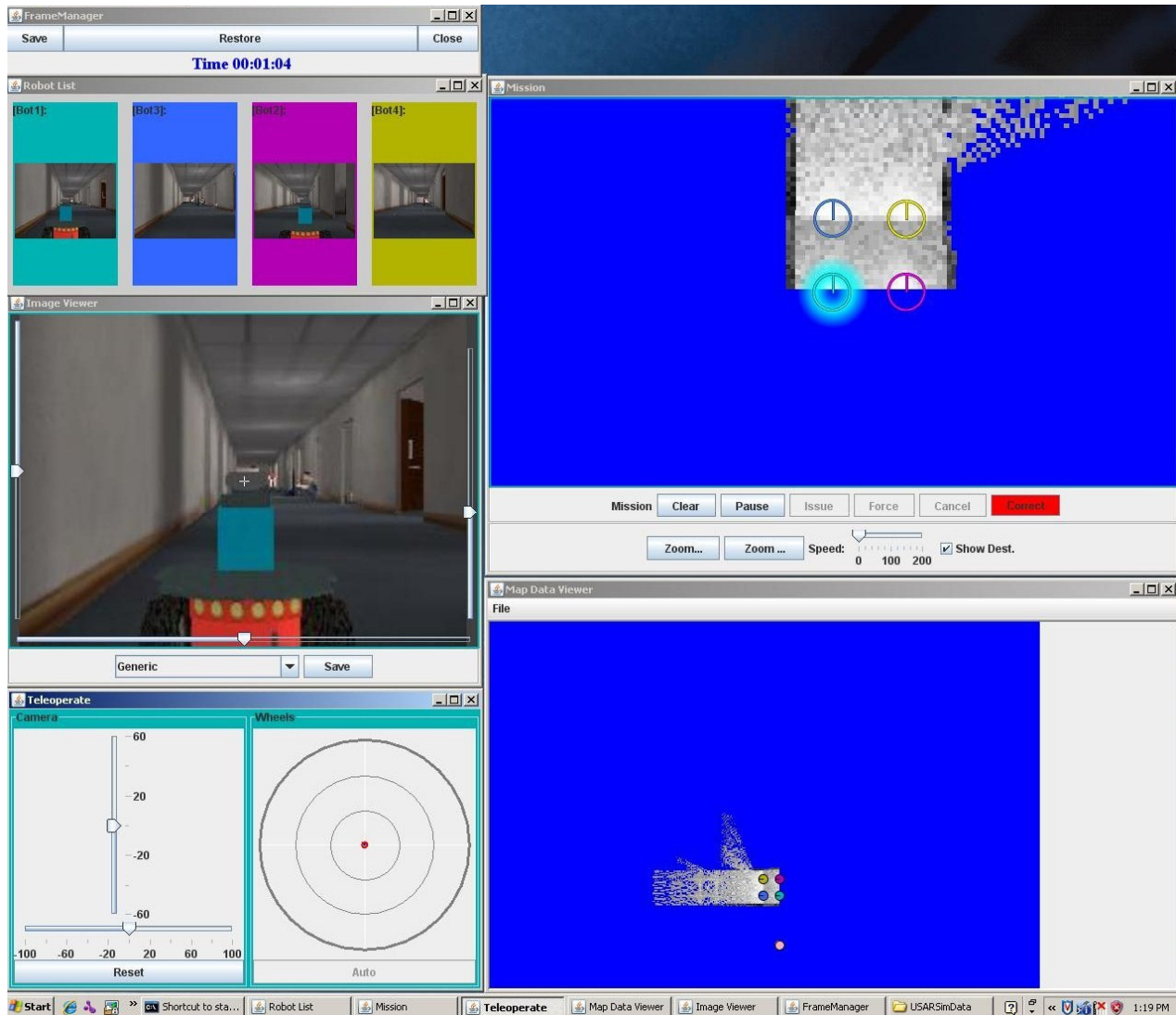


Figure 23. USARSim Human Interface.

UGV mission/display: In this condition, participants supervised four UGVs conducting a search mission for victims in a warehouse. The operators were responsible for directing the UGV's around the warehouse in search of victims (Figure 23). The objective of the mission was to find and mark the location of as many victims as possible. The Robot List (upper left) showed each robot's camera view, whereas the Image Viewer (middle left) displayed the selected robot's camera view. Robots could be controlled by either setting way-points on the Mission panel (upper right), or by teleoperation (lower left). The Mission panel (Appendix F) allowed operators to create, clear, and modify waypoints. Panoramic images were

taken at the terminal point of waypoint sequences, which were displayed on the Image Viewer (middle left) with operator's request. Through the teleoperation panel, it was also possible to pan and tilt the cameras. After victims were spotted in the UGV's video or in the panoramic images, operators were responsible for marking the victim's location on the Map Data Viewer (lower right frame) using a pop-up frame (Appendix F).

Dependent Variables

The main variable of interest for MCH-UVD validation was the MCH-UVD rating that the participants classified the display by, after completing the experimental trials. Operator performance and level of interface misuse were assessed by the test proctor on an observer rating scale (Appendix G). To capture perceived workload, participants were asked to rate how challenged they felt on a scale from 1 to 5 (Appendix H). The intent of this experiment was not to compare the two UV displays per se since they were inherently different tasks, but to compare how well MCH-UVD helped identify deficiencies in different displays. Thus, operator performance, level of interface misuse, and subjective workload were not compared across the two UV displays but were used to identify and control for the associations between these human behavior metrics and the MCH-UVD scale. In particular, the association between level of interface misuse and MCH-UVD scale was investigated to assess the construct validity of the scale.

Procedure

Participants first read and signed the Consent to Participate Form (Appendix I) in which they learned about the purpose of the experiment, compensation conditions, and the experimental aspects they would be asked to complete. Participants then filled out a brief demographic survey (Appendix J). Depending on the UVD condition assigned, the test proctor walked the participants through the OPS-USERS or USARSim display to illustrate the functionalities. The test proctor then demonstrated display functionalities and experimental tasks while the simulation was running. Participants were then given 12 minutes to practice with the display and ask questions. The test proctor then restarted the simulator and began to record performance metrics. The test scenario lasted 15 minutes. Following the test scenario, participants rated their perceived workload. Then, the test proctor showed the participants the General or Specific MCH-UVD, dependent on the assigned condition and gave a brief explanation of how to use the MCH-UVD. Participants then reported their MCH-UVD rating with an explanation for why they chose that rating and if they found any additional problems with the display (Appendix K). The participants were instructed to examine all the MCH-UVD ratings before indicating these additional problems. The test proctor then asked a few brief post-test questions (Appendix L) before compensating the participant.

EXPERIMENTAL RESULTS

MCH-UVD Ratings

As mentioned in the previous chapter, the scales evaluated in this experiment included “change blindness” rather than “poor task switching support” under rating seven. The experimental data showed that there were no responses in this rating level. The lack of responses for “change blindness” can be explained in part by its large overlap with rating eight, “poor attention direction”, and the term “poor attention direction” being easier to understand and thus potentially leading participants to select it rather than “change blindness”. In part due to these results, in the final scale presented in the earlier chapters, “change blindness” is replaced with “poor task switching support” to more clearly distinguish the display deficiencies represented in ratings seven and eight. The raw data for the MCH-UVD rating and the accompanying subjective responses can be found in Appendix M.

Statistical analyses were performed on the experimental data to compare MCH-UVD ratings across conditions. An ordered logit model, specifically proportional odds, was developed to compare the adequacy for the four general groupings of ratings (i.e., display is acceptable, deficiencies warrant improvement, deficiencies require improvement, and mandatory redesign) between different UVD (UAV and UGV) and MCH types (general and specific). In order to investigate and control for the effects of different operator characteristics and behavior on the MCH-UVD rating, the ordered logit model was adjusted for individual differences (age and gender), subjective workload, and experimenter ratings of the operator (operator competence, level of mission success, and level of interface misuse). In particular, the association between level of interface misuse and MCH-UVD scale was investigated to assess the construct validity of the scale.

Table 3. Number of observations in general MCH-UVD categories by experimental conditions.

UVD type	MCH type	MCH-UVD Rating			N
		Display is acceptable	Deficiencies warrant improvement	Deficiencies require improvement	
UAV	General	5	6	4	15
	Specific	1	8	6	15
UGV	General	2	8	5	15
	Specific	1	9	5	15

As shown in Table 3, none of the participants selected a rating of 1 or 2. Thus, participants did not think that a redesign was mandatory for these two displays. The statistical results showed that UVD type, MCH type, and UVD x MCH interaction were not significant. Thus, there was not enough statistical evidence to suggest a difference between the general and specific MCH scales for the two UV displays evaluated. However, in the general MCH setting, five participants rated the UAV display to be acceptable,

whereas in the specific MCH setting, there was only one participant who did the same. While not statistically significant, there is a trend that suggests that the more general scale could lead to more optimistic ratings.

Age, gender, subjective workload, operator competence and level of mission success were not significant ($\alpha=.05$), and hence were not included in the final model. The observer rating question assessing the level of interface misuse had four response alternatives: many, some, a few, and no misuses. Because there were only a few data points in each of these alternatives, the levels were combined and collapsed into two final levels: many or some misuses and few or no misuses.

The results of the statistical analysis revealed that the level of interface misuse had a significant effect ($\chi^2(1)=9.08, p=.003$) (Table 4). Compared to having many or some misuses, having a few or no misuses increased the odds of selecting a better rating by an estimated 516% (Odds-Ratio: 6.16, 95% CI: 1.89, 20.09). This provides evidence on the construct validity of the scale suggesting that the order of severity indicated in the MCH-UVD is directly proportional to the level of operator misuse. That is, the displays, which induce more operator misuses, are also placed on a higher severity rating in MCH-UVD.

Table 4. Number of observations in general MCH-UVD categories by level of misuse.

Level of Misuse	MCH-UVD Rating			N
	Display is acceptable	Deficiencies warrant improvement	Deficiencies require improvement	
Few or None	9	25	10	44
Some or Many	0	6	10	16

Display Deficiencies Identified

This section presents the participant feedback which accompanied their MCH-UVD ratings. The number of UV display deficiencies that were identified by each participant is compared across the general and specific scales. This analysis is then followed by a comparison on the total number of unique display deficiencies collectively identified by all the participants in each UVD type.

UAV display: Based on Figure 24, it appeared that individually, participants were likely to identify more UAV display deficiencies with the specific scale compared to the general scale. This effect was only marginally different between the general and specific scales ($\chi^2(1)=3.13, p=.08$).

A large number of deficiencies identified by all the participants in the general scale condition were identical to the deficiencies identified by all the participants in the specific scale condition. Four participants (27%) in each MCH condition identified a lack of display support for search guidance. A few participants indicated that better information on the remaining fuel level was necessary (general: 2 participants, specific: 1 participant). Participants also identified deficiencies related to path planning and

re-planning. In particular, the participants indicated that time delays with respect to control inputs, automatic updating, and accepting and prioritizing tasks made it harder to use the interface (general:5 participants, specific: 7 participants). One participant pointed out that the automated plan suggested merely mirror the route input by the operator, thus, creating unnecessary time delays. Four participants (27%) in each MCH condition indicated that the imprecise UAV paths were a problem. Participants indicated that it was difficult to change the UAV flight path (general: 3 participants, specific: 4 participants). Additional waypoint manipulation capabilities were suggested to overcome the problems related to path-planning.

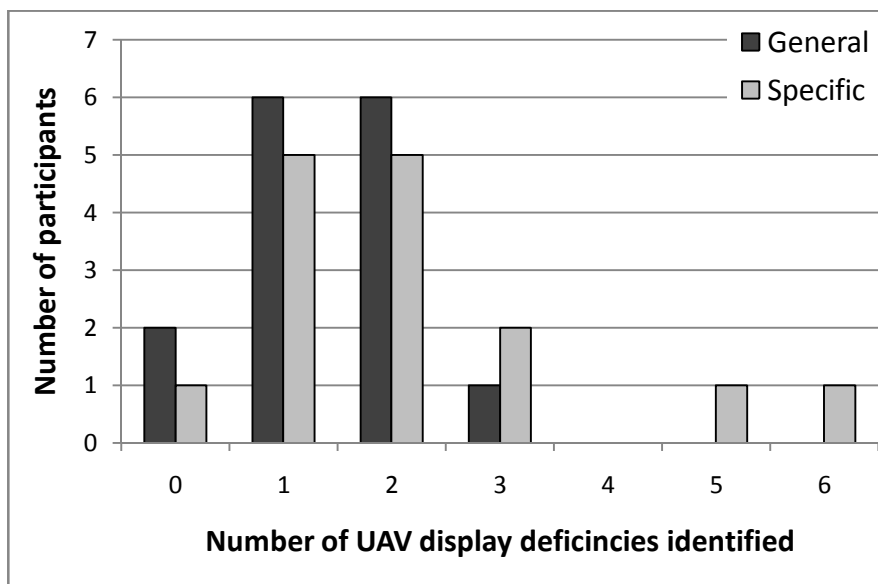


Figure 24. Number of UAV display deficiencies identified.

Unique deficiencies identified. Only one unique display deficiency was identified with the General scale, which was the lack of information provided on the consequences of selected actions (1 participant). There were a total of four deficiencies uniquely identified with the specific scale. These deficiencies included the lack of information on UAV flight parameters (i.e., direction, speed), obscured duration settings, difficult target detection, and the large number of steps required to change the search patterns as well as to track targets (each by 1 participant).

UGV display: The number of identified UGV display deficiencies was not statistically significantly different between general and specific scales (Figure 25).

Similar to the UAV display, a large number of deficiencies identified for the UGV display with the general scale were identical to the deficiencies identified with the specific scale. Out of the 15 participants in each MCH condition, the majority identified time delays to be problematic, especially the delays associated with manual control of the vehicles and the cameras, as well as the slow UV movement in

general (general: 10 participants, specific: 7 participants). It was suggested to have multiple cameras on a vehicle to avoid rotating the camera. It was also indicated by one participant in each MCH type that the robots did not always move smoothly and follow waypoints exactly. One participant in each condition indicated that the additional step of clearing UGV paths was difficult and unnecessary. Display not indicating the obstacles on UGVs' paths was another deficiency identified with both the general (2 participants) and specific scales (4 participants). Participants considered the two maps being confusing and unnecessary (1 participant in each condition). Clutter was also deemed to be a problem (general: 1 participant, specific: 2 participants).

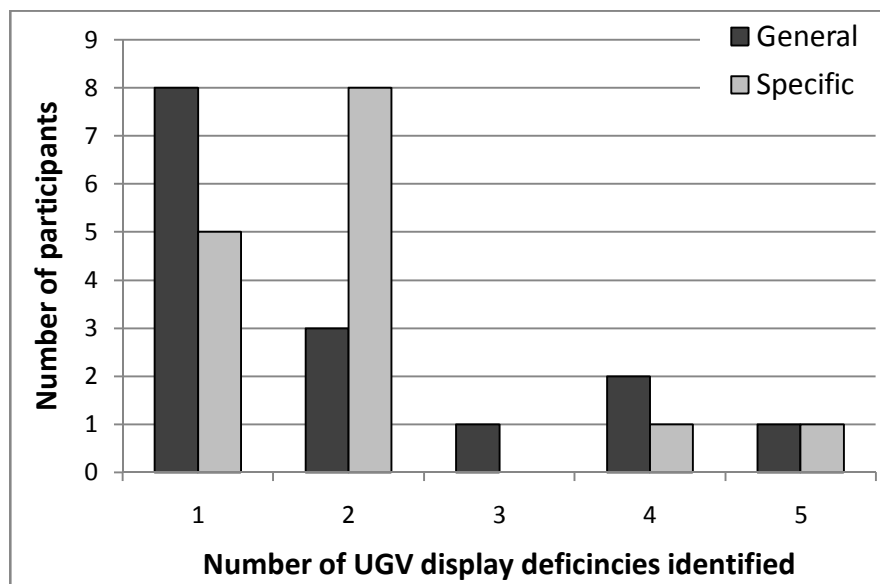


Figure 25. Number UGV display deficiencies identified.

Unique deficiencies identified. There were a total of six deficiencies uniquely identified with the general scale. These problems included the blue background and the pop-up distractions (1 participant each), the lack of alerts before two UGVs collided (1 participant), the lack of indication when UVs were idle (1 participant), and the lack of display customizability (1 participant). Suggestion to improve the display consisted of physically coloring the UGVs to support easier identification of robots (1 participant).

There were three uniquely identified deficiencies with the specific scale. These deficiencies included the difficulty in switching between the four robots views (1 participant), as well as the issues related to UGV orientation and depth perception. Specifically, the camera angle made it difficult to know how UGVs were oriented (3 participants), and it was hard for participants to estimate distances on the map based on the video feed depth perception (6 participants). Therefore, participants had to place several markers to get to a desired location. This display deficiency (inaccurate goal assignment) was identified by a large number of participants with the Specific scale and by no participants with the General scale, and is critical since it can significantly interfere with UV control. Thus in this case, the Specific scale

helped a larger number of operators identify a major display deficiency, which was not captured by the General scale.

Feedback on MCH-UVD

This section presents the participant responses to post test debrief questions. The raw data can be found in Appendix N.

Support in identifying deficiencies: Participants were asked if MCH-UVD helped them identify deficiencies. The majority of participants responded yes to this question. There were a total of three participants in the UAV condition who did not identify any display deficiencies. Out of the 57 participants who identified deficiencies, 49, that is, 86% (23 in general, 26 in specific) thought that MCH-UVD helped them identify these display problems. 32% (10 in general, 8 in specific) of the 57 participants said that they could not have recognized these deficiencies without the help of MCH-UVD. 14% (4 in general, 4 in specific) said that they could have identified deficiencies but would not be able to indicate the severity. An additional 12% (5 in general, 2 in specific) also indicated that they could have identified deficiencies but would not be able to describe them accurately.

Participants were also asked if MCH-UVD helped them identify all deficiencies. The responses suggested that some of the participants considered this question as referring to the single rating they selected rather than the whole scale. If there were any other deficiencies, which belonged to different rating levels other than the selected one, these participants responded “no” to the question. Therefore, the responses to this question are not analyzed due to possible confounding. It is important to note that some participants identified problems associated with system functionality (e.g., slow UV movement, imprecise UAV path, large time lags) and manual control of robots, which were not covered by MCH-UVD. In general, while evaluating the display, participants tended to point out these problems in addition to display deficiencies.

MCH-UVD design: There were mixed responses with respect to the design of the scales. The aspects of MCH-UVD categorized as being most useful included the detailed descriptions of display issues (general: 7 responses, specific: 14 responses), flowchart (general: 16 responses, specific: 11 responses), severity scale (general: 1 response, specific: 2 responses), and color coding (general: 2 responses, specific: 4 responses). The aspects of MCH-UVD categorized as being least useful included the flowchart (general: 9 responses, specific: 10 responses), technical terms (general: 13 responses, specific: 13 responses), and wording being too long (general: 3 response, specific: 16 responses).

Overall, the views on the usefulness of the flowchart format were split about in half. Some participants categorized the flowchart to be the most useful aspect guiding them in their ratings, whereas

others thought that the flowchart questions were too broad, and led them to the wrong ratings. 9 participants (15%) suggested using checklists rather than picking one specific rating.

43% of the participants (half in general, half in specific) found the language to be too technical and difficult to understand at times, and 32% (general: 6%, specific: 26%) found the wording to be too long. 23% of the participants (general: 6%, specific: 17%) suggested having more rankings on the scale including more display issues.

DISCUSSION

The results of the validation experiment revealed that MCH-UVD has good construct validity as suggested by the significant relation observed between the level of interface misuse and the MCH-UVD rating. Participants who had more misuses with a display gave it a worse rating. Moreover, almost all of the participants (86%) thought that MCH-UVD helped them identify display deficiencies, and 32% of the participants said that they could not have identified the deficiencies otherwise. Thus, the MCH-UVD can guide operators in display evaluation by presenting the type of display characteristics they should be looking for.

The experiment compared the specific and the general scales for two types of UVDs: UAV and UGV. It should be noted that, although current work focused on Specific MCH-UVD diagnosis tool development for UAVs and UGVs, scales can also be developed for other unmanned vehicles such as underwater unmanned vehicles (UUVs) or space unmanned vehicles. Overall, there was not any statistical evidence to suggest a difference between the ratings obtained with the specific and general scales for the two UV displays tested. Although only marginally statistically significant, individuals appeared to identify more deficiencies with the specific scale as compared to the general one. Given the limited experimental sample size, this is an important finding which has implications for the use of MCH-UVD in practice. When there are only a few operators available to rate a UV display, these results show that more deficiencies can be identified with the Specific MCH-UVD.

There were a total of three and six uniquely identified deficiencies in the Specific and General scales, respectively. While more unique deficiencies were identified with the General scale, these occurrences were not clustered around any clear problems, while those unique deficiencies identified with the Specific scale were clustered around a major design flaw not identified with the General scale. Thus, the likelihood that the Specific scale could identify a major display deficiency is higher than with the General scale. Longitudinal data from actual practice when the scales are in use could provide more insight on how much additional benefit the specific scale provides and if this additional benefit is worth the effort. The amount of time required generating the Specific scale and the additional benefit it may provide creates a trade-off for the practitioners. Although, the method to generate Specific scales is relatively straightforward, specific MCH-UVD scales may need frequent updating with changes in mission

requirements or specifications, whereas the general scale can provide significantly faster feedback during the rapid prototyping of displays for different mission requirements.

Another disadvantage of Specific MCH-UVD diagnosis tools is that descriptions of operator demands are sometimes wordy due to inclusion of unique mission details. These long descriptions can be tedious for operators to read, causing them to potentially gloss over or fixate upon a certain part of the scale. The subjective feedback from the experiment revealed that even the general scale was found to be wordy with confusing technical language. However, the level of technicality, and the number of words needed to identify technical terms is a tradeoff, which has to be decided upon by the practitioners. In fact, 47% of the participants who used the Specific scale indicated that the detailed descriptions were the most helpful aspect of MCH-UVD, as opposed to 23% of the participants who used the General scale. If the operators are familiar with the technical terms, the amount of words used to explain them can be reduced. In contrast, if the operators are not familiar with the technical words, longer and more detailed descriptions would be required.

In current operations, UVs rarely only serve one purpose and often are designed to perform multiple missions, either singly or concurrently. An advantage of the Specific MCH-UVD diagnosis tool is that it can be custom designed to consider more than one mission in the decision diamonds and operator demands descriptions. Display developers and testers can either choose to administer multiple single-mission Specific MCH-UVD scales for each mission type, or combine the information of multiple missions into one Specific MCH-UVD scale to see how a display supports all missions at once. While the former method may better pinpoint display deficiencies pertaining to one or more of a UV's missions, the latter method is a faster way to evaluate a multi-mission UV display. Examination of how to represent multiple UV missions through one or many Specific MCH-UVDs is an open question.

Although both the General and Specific MCH-UVD scales serve as useful tools for evaluating displays and diagnosing their problems, they do have limitations. Both MCH-UVD scales should be thought of as one tool in a large toolkit of methods to evaluate displays, particularly because the scales are quasi-objective. Additionally, the scales are not comprehensive as an evaluation methodology because both scales have limited descriptions due to their design as quick subjective assessment tools. The goal behind the MCH-UVD diagnosis tool is to efficiently gather an operator's display assessment through a standardized quasi-subjective scale, linking judgments to specific display problems and limitations. Because MCH-UVD diagnosis tools only provide one subjective measure of operator-UV interaction, other objective metrics need to be collected to get an overall picture of how a UV display supports an operator in supervising a mission.

How to Administer MCH-UVD

While this technical report has detailed the General MCH-UVD diagnosis tool and how to adapt it into a Specific MCH-UVD scale, little has been said about administering the tool. The MCH-UVD scales provide an inexpensive and easy way to determine required design improvements, as display developers can receive quick feedback about a display. MCH-UVD scales can be used throughout the display design process in an iterative manner to ensure a final display design that fully supports information processing stages. The MCH-UVD scales can even be administered with initial developmental mock-up displays to evaluate support of safely completing a mission and proposed decision support tools

The MCH-UVD tool should be administered immediately after an operator has used a UV display. It can be provided to the operator simply on a sheet of paper or on a computer screen. In the validation experiment, after examining all the MCH-UVD ratings, 60% of the participants identified additional deficiencies which were not covered by the specific rating they selected. Thus, after UV operators rate a display, they should provide written comments explaining why they chose a particular rating so as to pinpoint the exact display deficiencies, but also should be asked about other possible deficiencies.

CONCLUSION

The simple, flexible, and quasi-subjective nature of the MCH-UVD diagnosis tool make it well poised to become a useful metric in UV display evaluation. The experiment conducted to validate MCH-UVD revealed that when the operators are presented with this tool, they can identify display deficiencies that they cannot identify otherwise. Thus, the MCH-UVD can guide the operators to look for certain display characteristics required to support UV missions.

By using an intuitive questioning method, the General MCH-UVD scale can be adapted to create a UV and mission-targeted Specific MCH-UVD scale. Most of a Specific MCH-UVD scale mirrors the General scale, but includes extra details and definitions pertaining to a specific UV and mission. While the General scale can be universally used to uncover display deficiencies, the Specific scale can determine what aspects of a mission the UV display is not supporting. However, Specific scale use requires frequent updating with changes in mission requirements or specifications, with the cost of updating possibly surpassing the potential benefits. In fact, the validation experiment did not reveal substantial benefits with the use of the Specific scale. Therefore, the General MCH-UVD appears to be a promising tool, which can provide rapid prototyping feedback and help operators rate different UV displays.

ACKNOWLEDGEMENTS

This research is funded by the United States Army Aberdeen Test Center. The views and conclusions presented in this paper are those of the authors and do not represent an official opinion, expressed or implied, of the United States Army. Special thanks are extended to Ryne Barry and Javier

Garcia, the undergraduate assistants who helped with data collection and processing. We would also like to thank the following persons for providing their unmanned vehicle displays and support: Olivier Toupet from Aurora Flight Sciences, Paul Scerri and Prasanna Velagapudi from Carnegie Mellon University, Michael Lewis from the University of Pittsburgh, and Cameron Fraser from MIT.

REFERENCES

- Barbato, G., Feitshans, G., Williams, R., & Hughes, T. (2003). *Unmanned combat air vehicle control and displays for suppression of enemy air defenses*. Paper presented at the 12th International Symposium on Aviation Psychology, Dayton, OH.
- Borchers, P. F., Franklin, J. A., & Fletcher, J. W. (1998). *Flight research at Ames, 1940-1997: fifty-seven years of development and validation of aeronautical technology (NASA SP-3300)*. Moffett Field, CA: NASA, Ames Research Center.
- Bredereke, J., & Lankenau, A. (2002). A rigorous view of mode confusion *Lecture Notes in Computer Science*. Berlin: Springer.
- Carrigan, G. P., Long, D., Cummings, M. L., & Duffner, J. (2008). Human factors analysis of predator B crash *Proceedings of AUUSI: Unmanned Systems North America*. San Diego, CA.
- Cooper, G. E. (1957). Understanding and interpreting pilot opinion. *Aeronautics Engineering Review*, 16(3), 47-51.
- Culbertson, E. (2006). COMUSAFE: Unmanned Aircraft Key to Future Decision Superiority Retrieved March 18, 2008, from <http://www.af.mil/news/story.asp?id=123029520>
- Cummings, M. L., Myers, K., & Scott, S. D. (2006). Modified Cooper Harper evaluation tool for unmanned vehicle displays *Proceedings of UVS Canada: Conference on Unmanned Vehicle Systems Canada*. Montebello, PQ, Canada.
- Dixon, S., Wickens, C. D., & Chang, D. (2005). Mission control of multiple unmanned aerial vehicles: a workload analysis. *Human Factors*, 47, 479-487.
- GlobalSecurity.org. Military robots/unmanned ground vehicles (UGV) Retrieved January, 7, 2007, from <http://www.globalsecurity.org/military/systems/ground/ugv.htm>
- Harper, R. P., & Cooper, G. E. (1986). Handling qualities and pilot evaluation. *Journal of Guidance, Control, and Dynamics*, 9(5), 515-529.
- Hart, S. G., & Staveland, L. E. (1988). The subjective workload assessment technique: a scaling procedure for measuring mental workload. In P. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam, The Netherlands: North Holland B. V.
- Lewis, M., Wang, J., & Hughes, S. (2007). USARsim: Simulation for the study of human-robot interaction. *Journal of Cognitive Engineering and Decision Making*, 1(1), 98-120.

- McFarlane, D. C. (2002). Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction*, 17(1), 63-139.
- Nehme, C., Crandall, J. W., & Cummings, M. L. (2007). *An operator function taxonomy for unmanned aerial vehicle missions*. Paper presented at the 12th International Command and Control Research and Technology Symposium, Newport, RI.
- NTSB (1995). *Marine Accident Report, Grounding of the Panamian Passenger Ship Royal Majesty on Rose and Crown Shoal near Nantucket, Massachusetts* (No. NTSB/MAR-97/01): National Transportation Safety Board.
- Parasuraman, R., Sheridan, T., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans*, 30(3), 286-297.
- Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124(2), 207-231.
- Rubinstein, T., & Mason, A. F. (1979). The accident that shouldn't have happened: an analysis of the Three Mile Island. *IEEE Spectrum*, November, 33-57.
- Sheridan, T. B. (1972). On how often the operator supervisor should sample. *IEEE Transactions on Systems, Science, and Cybernetics*, SSC-6, 140-145.
- Tsach, S., Peled, A., Penn, D., Keshales, B., & Guedj, R. (2007). *Development trends for next generation UAV systems*. Paper presented at the AIAA 2007 Conference and Exhibit, Rohnert Park, California.
- Wang, J., & Lewis, M. (2007). Human control of cooperating robot teams *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. Arlington, VA.

APPENDICES

- A:** UGV Taxonomy Mission Descriptions
- B:** General MCH-UVD Tested in the Validation Experiment
- C:** Specific UAV Target Acquisition MCH-UVD Tested in the Validation Experiment
- D:** Specific UGV Search MCH-UVD Tested in the Validation Experiment
- E:** UAV Display: Accepting Automation's Plan and Tracking Task
- F:** USARSim Mission Panel and Map Data Viewer
- G:** Observer Rating Scale
- H:** Subjective Workload Rating
- I:** Consent to Participate
- J:** Demographic Survey
- K:** Evaluation From for MCH-UVD Rating
- L:** Post Test Debrief
- M:** MCH-UVD Rating Data
- N:** Post Test Debrief Responses

Appendix A: UGV Taxonomy Mission Descriptions

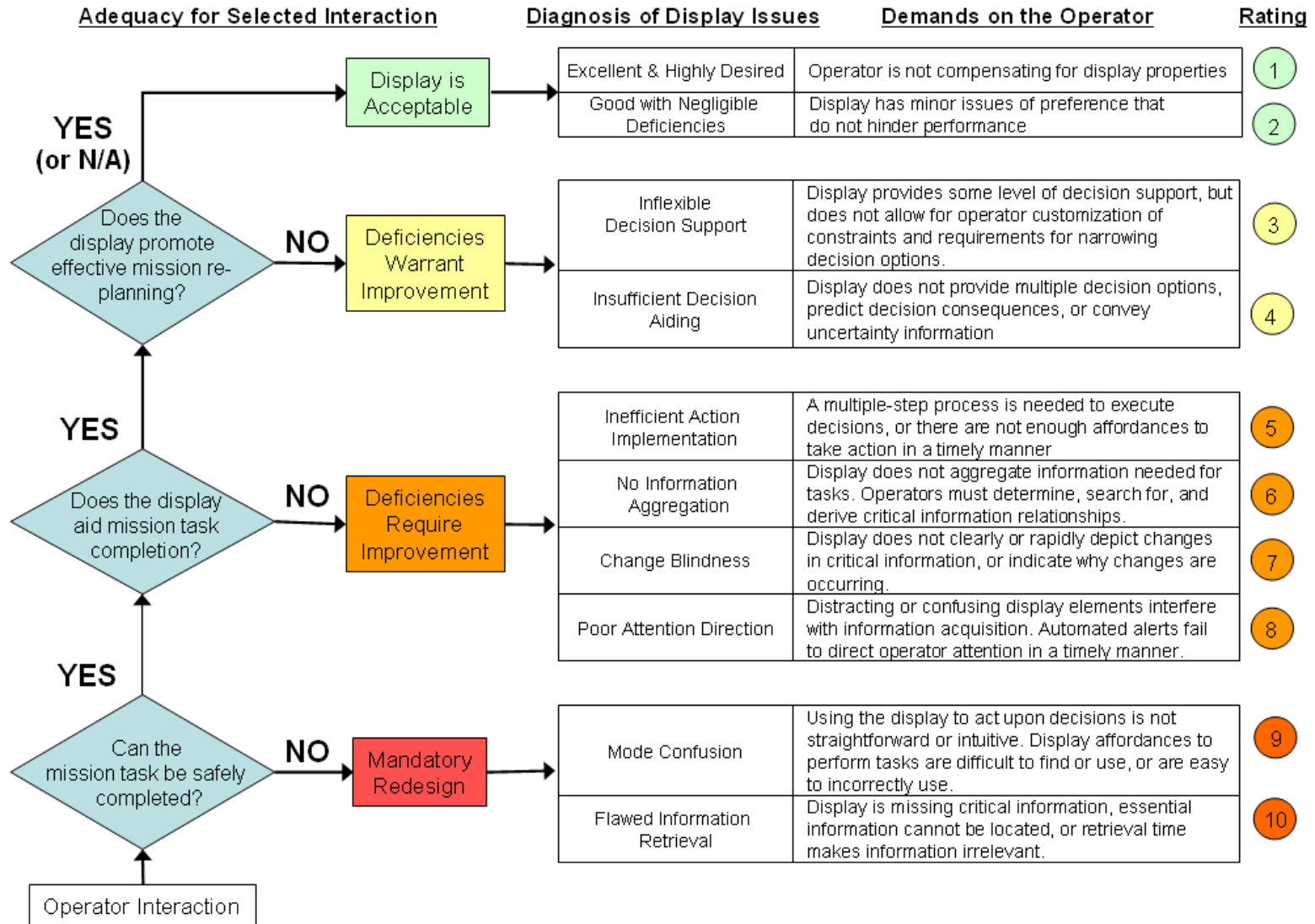
The UGV Mission Taxonomy was created for use in development of vehicle and mission-specific MCH-UVD diagnosis tools. This taxonomy represents UGV missions that are performed by UGVs today, as well as missions that could be performed in the not-to-distant future. This appendix provides definitions of missions included in the UGV Taxonomy presented in Section 0, and gives examples of current UGVs performing these missions from (GlobalSecurity.org).

- **Logistics** is the procurement, supply, maintenance, support, and transport of equipment or personnel, and comes in four different missions:
 - **Construction** is the act of building structures or supporting the building of structures. Examples: Automated Ordnance Excavator (AOE), All-purpose Remote Transport System (ARTS)
 - **Transport** is the movement of physical items, such as *passengers* or *cargo*. Examples: Mesa Associates' Tactical Integrated Light-Force Deployment Assembly (MATILDA), Multifunction Utility/Logistics Equipment Vehicle (MULE UGV), All-purpose Remote Transport System (ARTS)
 - **Water Purification & Generation** is the act of filtering and decontaminating water, and/or creating water from constituent hydrogen and oxygen. Examples: Multifunction Utility/Logistics Equipment Vehicle (MULE UGV)
 - **UAV Launch & Landing** is the act of physically supporting UAV launch and landing with the UGV. Examples: Light Armored Vehicle (LAV II)
- **Payload Delivery** refers to the delivery of weapons for attack, and comes in two forms:
 - **Lethal** payload could be a missile or other device meant to kill. Examples: TALON; Gladiator Tactical Unmanned Ground Vehicle; Remote Detection, Challenge, and Response System (REDCAR); Armed Robotic Vehicle (ARV) UGV;
 - **Non-Lethal** payload is something to debilitate or impede, but not kill, an adversary. Fog and smoke are examples of non-lethal payloads. Examples: Combined Operations Battlefield Robotic Asset (COBRA), TALON, Gladiator Tactical Unmanned Ground Vehicle, Remote Detection, Challenge, and Response System (REDCAR)
- **Decoy** refers to UGV usage as a lure or for deception.
- **Intelligence, Reconnaissance, Surveillance, and Sensing** is the search and gathering of useful information and the monitoring of people, objects, or processes. The UGV Mission Hierarchy contains a variety of sub-missions under this classification:
 - **Battle Damage Assessment (BDA)** is the act of measuring the status of a target. Examples: Armed Robotic Vehicle (ARV) UGV
 - **Environment** sensing examines aspects of the environment and can be in the form of *nuclear, biological, chemical, and other*. Example: Combined Operations Battlefield

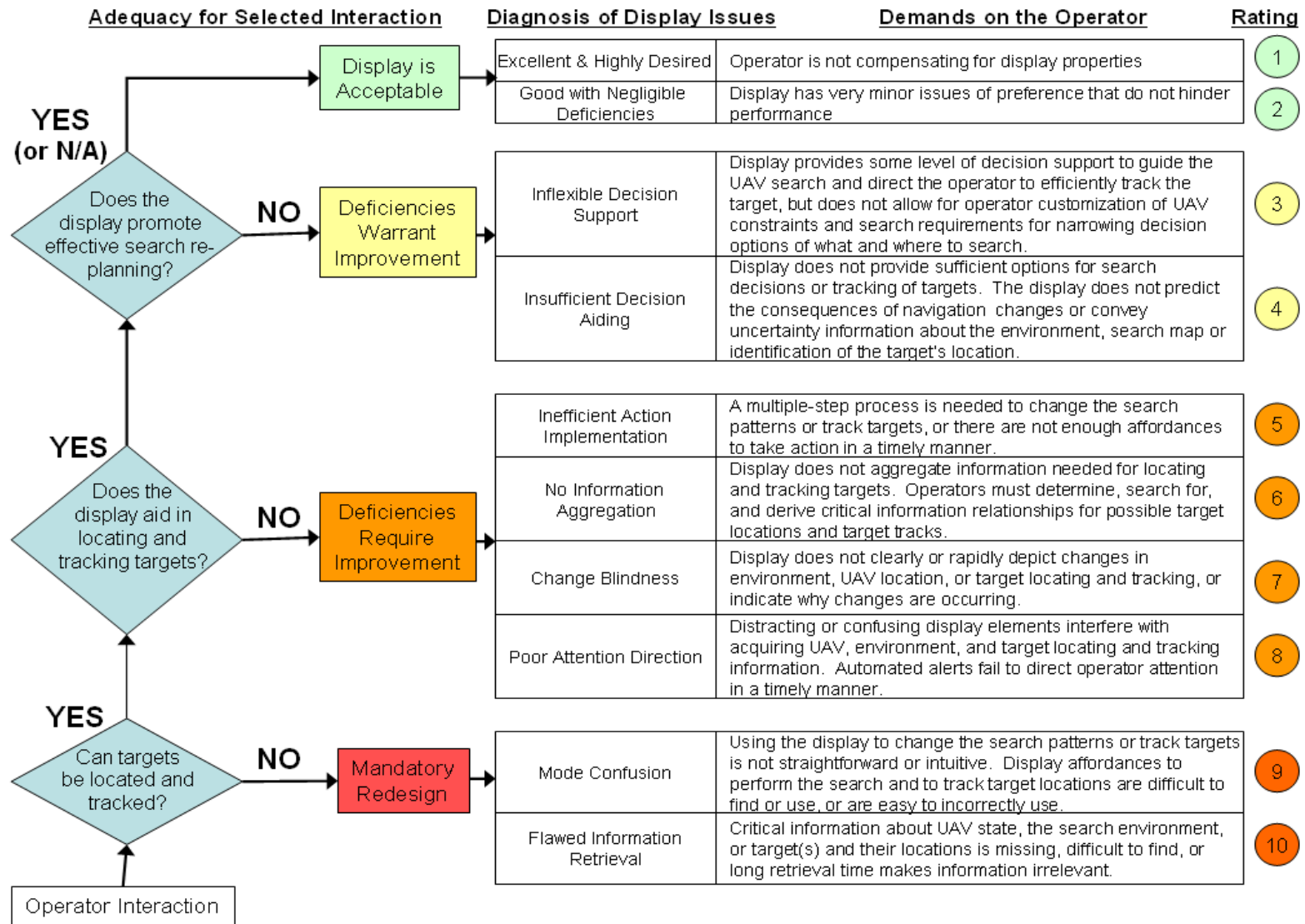
- Robotic Asset (COBRA); Dragon Runner; Omni-Directional Inspection System (ODIS); Packbot; TALON; Multifunction Utility/Logistics Equipment Vehicle (MULE UGV); Gladiator Tactical Unmanned Ground Vehicle, Mobile Detection Assessment and Response System (MDARS); Remote Detection, Challenge, and Response System (REDCAR); Surveillance and Reconnaissance Ground Equipment (SARGE)
- **Target Acquisition** is the detection of a target of interest, either *static* or *dynamic*. Examples: Mesa Associates' Tactical Integrated Light-Force Deployment Assembly (MATILDA); Gladiator Tactical Unmanned Ground Vehicle; Remote Detection, Challenge, and Response System (REDCAR)
 - **Listening** is the sensing of communication passing between other parties.
 - **Structural Integrity** sensing is the observation and evaluation of structures and their stability status. Example: Soldier UGV (SUGV)
 - **Object Detection** refers to the sensing of *minefields*, *obstacles*, *weapons*, or *inventory*. Example: Combined Operations Battlefield Robotic Asset (COBRA), Man Portable Robotic System (MPRS), Packbot, TALON, Mobile Detection Assessment and Response System (MDARS), Remote Ordnance Neutralization System (RONS), Armed Robotic Vehicle (ARV) UGV
 - **Mapping** is the act of creating a representation of an area in terms of physical forms and locations. Examples: Man Portable Robotic System (MPRS) Multifunction Utility/Logistics Equipment Vehicle (MULE UGV)
 - **Target Designation** is the act of marking or pointing out a target. Examples: TALON, Gladiator Tactical Unmanned Ground Vehicle
 - **Search & Rescue** finds and recovers missing animals or people. Examples: TALON
 - **Communication** is the conveyance of information between two parties, and can either be in the form of:
 - **Voice**. Examples: Remote Detection, Challenge, and Response System (REDCAR)
 - **Data**. Examples: Multifunction Utility/Logistics Equipment Vehicle (MULE UGV)
 - **Neutralization** is the counteraction or nullification of an object or entity, and pertains to a number of sub-missions:
 - **Fire** neutralization is the act of putting out a fire. Examples: All-purpose Remote Transport System (ARTS)
 - **Explosives** neutralization is the safe disposal of unsafe, explosive material. Examples of this could be defusing a bomb, or exploding a bomb in a safe place. Examples: Man Transportable Robotic System (MTRS), TALON, Mini-Flail, Robotic Combat Support System (RCCS), Remote Ordnance Neutralization System (RONS), All-purpose Remote Transport System (ARTS), Automated Ordnance Excavator (AOE)

- **Electronic** neutralization involves the use of electronic signals to *attack* other signals, or to *protect* one's own signals. Examples: Robotic Combat Support System (RCSS)
- **Obstacle** neutralization is the act of breaching or disarming a structure, obstacle, or trap. Example: Combined Operations Battlefield Robotic Asset (COBRA), Mesa Associates' Tactical Integrated Light-Force Deployment Assembly (MATILDA), TALON, Gladiator Tactical Unmanned Ground Vehicle, Robotic Combat Support System (RCCS), Armed Robotic Vehicle (ARV) UGV
- **Minefield** neutralization is the act of *path-clearing* or *de-mining* a minefield. Example: Combined Operations Battlefield Robotic Asset (COBRA), TALON, Multifunction Utility/Logistics Equipment Vehicle (MULE UGV), Mini-Flail, Robotic Combat Support System (RCCS), All-purpose Remote Transport System (ARTS), Automated Ordnance Excavator (AOE), M60 Panther, M1 Abrams Panther II (in teleoperation mode).

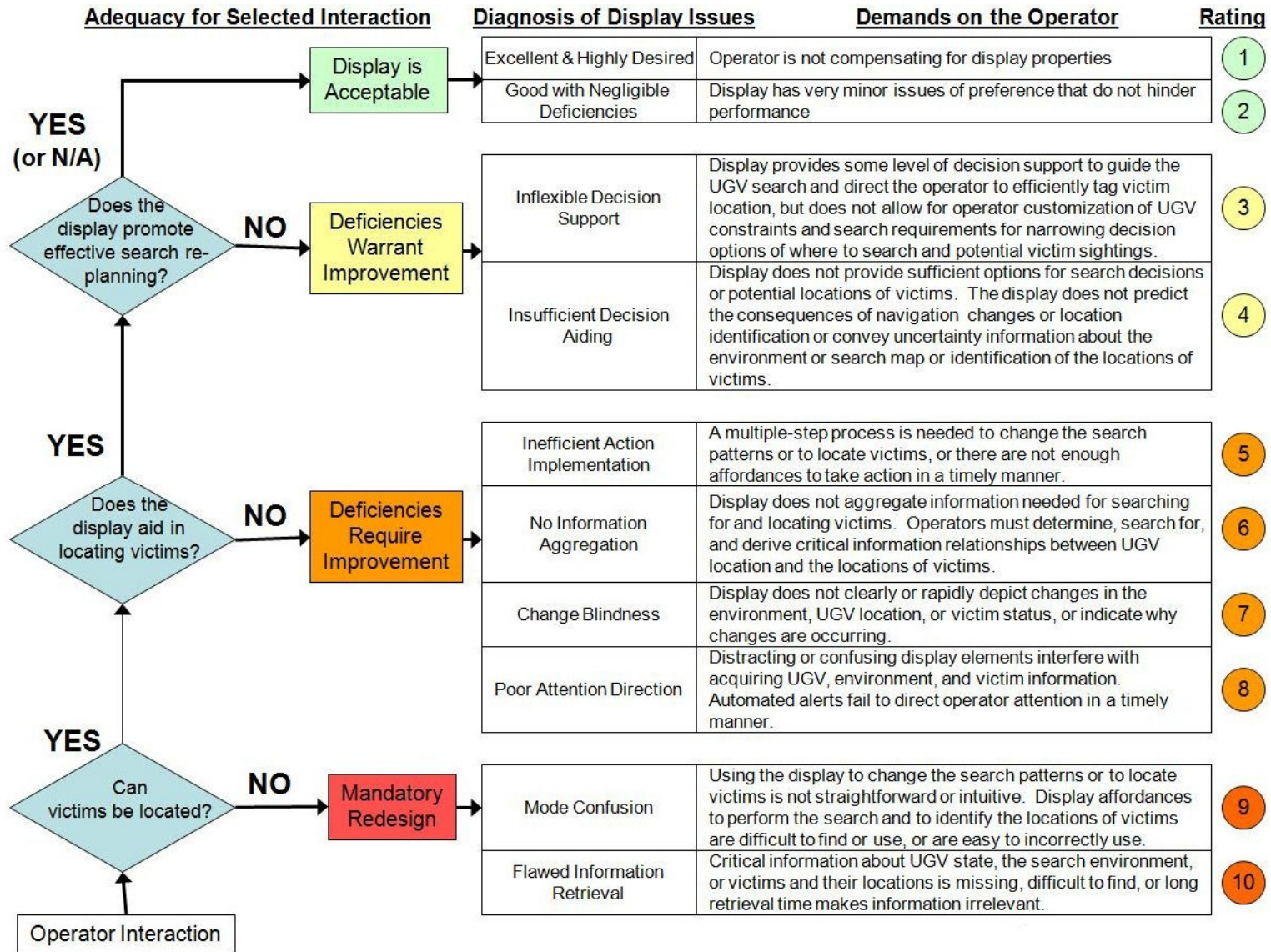
Appendix B: General MCH-UVD Tested in the Validation Experiment



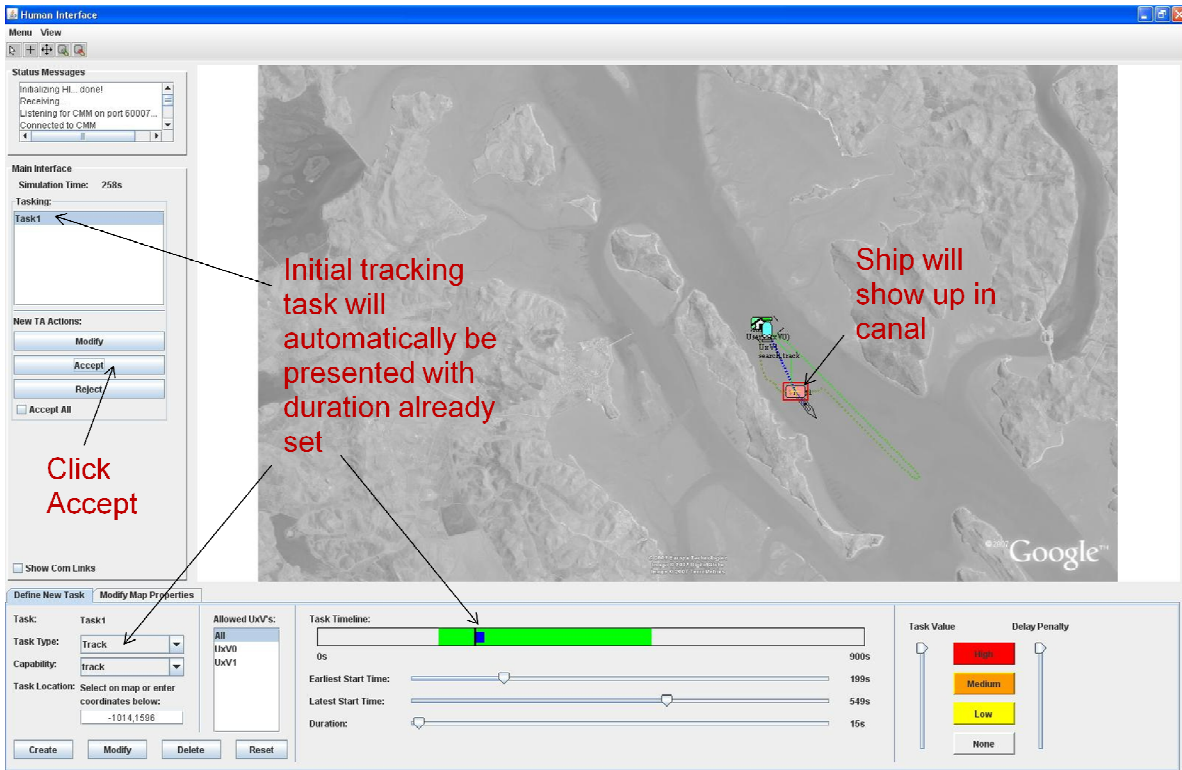
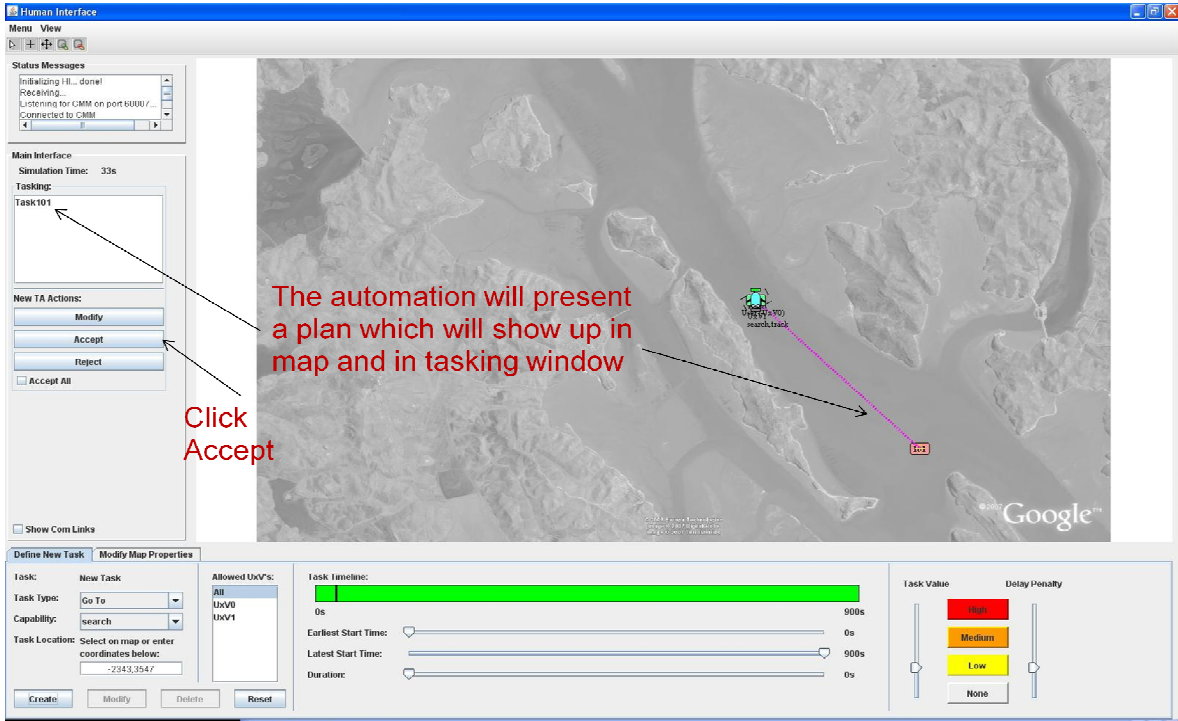
Appendix C: Specific UAV Target Acquisition MCH-UVD Tested in the Validation Experiment



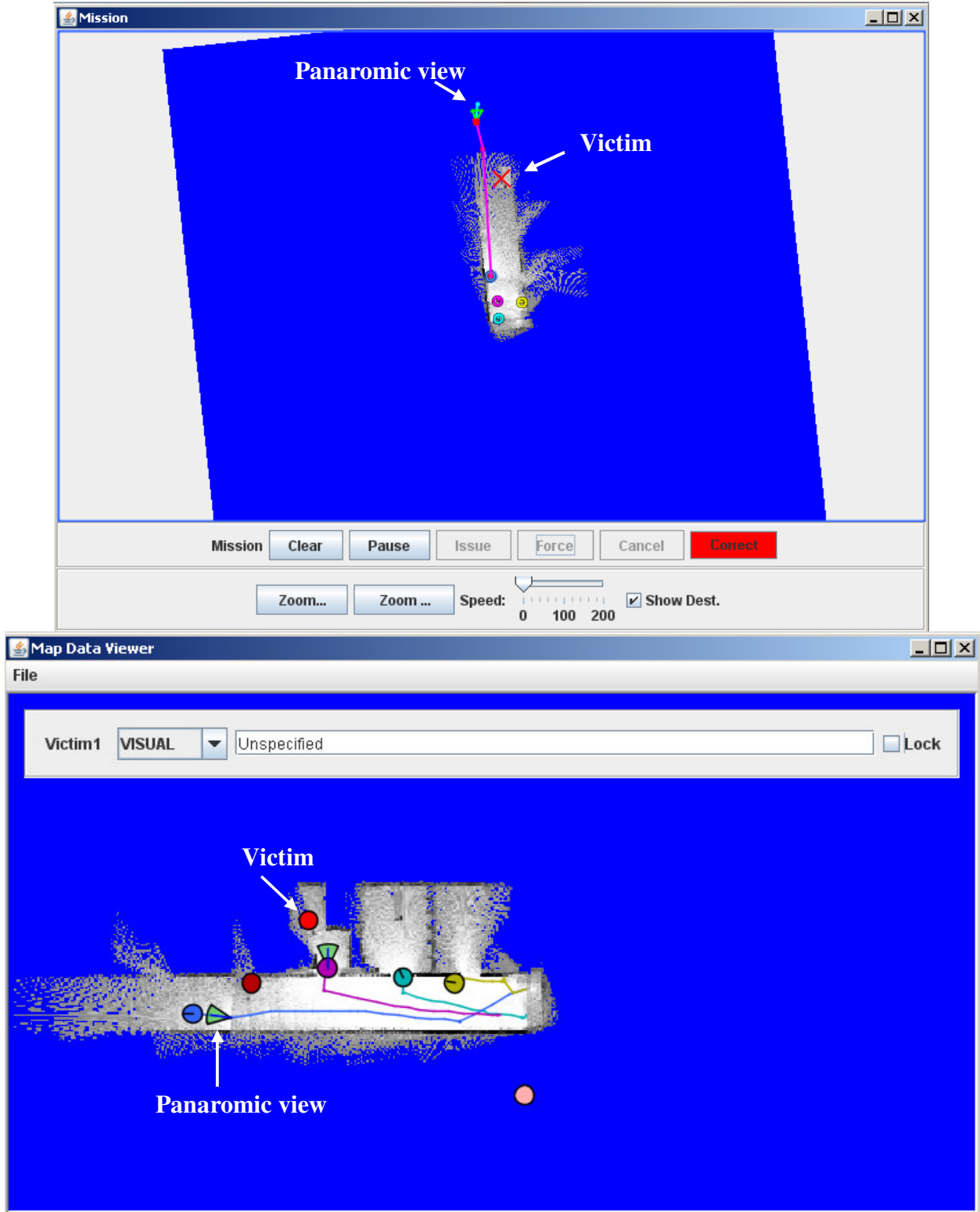
Appendix D: Specific UGV Search MCH-UVD Tested in the Validation Experiment



Appendix E: UAV Display: Accepting Automation's Plan and Tracking Task



Appendix F: USARSim Mission Panel and Map Data Viewer



Appendix G: Observer Rating Scale

OBSERVER RATING SCALE (UAV)

Interface Interaction – How well did the operator interact with the interface? (i.e., how much the operator misused the interface)

- 1 – Less than adequately (many misuses)
- 2 – Adequately (some misuses)
- 3 – Well (a few misuses)
- 4 – Very well (no misuses)

Operator Competence – Did the operator have a strategy and how well did he/she execute the strategy?

- 1 – Had no strategy
- 2 – Had a strategy but failed to execute the strategy
- 3 – Had a strategy and marginally executed the strategy
- 4 – Had a strategy and executed the strategy

Mission Completion

Searching – How well did the operator search the canal for ships? (i.e., were there ships that slipped by)

- 1 – Focused on small region of canal
- 2 – Covered half of the canal
- 3 – Covered the whole canal marginally
- 4 – Covered the whole canal well

Tracking – How well did the operator continue tracking identified targets? (were identified targets kept up with or were they lost)

- 1 – Not tracked at all after initial identification
- 2 – Tracked only after initial identification
- 3 – Tracked continually after identification
- 4 – Tracked continually while continuing search

Notes: (i.e., about operator strategies in supervising the vehicles, search pattern or automation reliance, as well as, abnormalities in the testing, etc...)

Number of Ships Located: _____ Number of Land Targets Located: _____

OBSERVER RATING SCALE (UGV)

Interface Interaction – How well did the operator interact with the interface? (i.e., how much the operator misused the interface)

- 1 – Less than adequately (many misuses)
- 2 – Adequately (some misuses)
- 3 – Well (a few misuses)
- 4 – Very well (no misuses)

Asset Implementation – How well did the operator use available assets? (were vehicles left unsupervised, note a vehicle that is not used is unsupervised)

- 1 – Only one vehicle was supervised
- 2 – Two vehicles were supervised
- 3 – Three vehicles were supervised
- 4 – All four vehicles were supervised

Operator Competence – Did the operator have a strategy and how well did he/she execute the strategy?

- 1 – Had no strategy
- 2 – Had a strategy but failed to execute the strategy
- 3 – Had a strategy and marginally executed the strategy
- 4 – Had a strategy and executed the strategy

Mission Completion – How well did the operator complete the mission? (i.e., were the victim labelings accurate or did the operator keep up with identified targets)

- 1 – Very poorly
- 2 – Poorly
- 3 – Well
- 4 – Very well

Notes: (i.e., about operator strategies in supervising the vehicles, search pattern or automation reliance, as well as, abnormalities in the testing, etc...)

Number of Victims Located: _____

Appendix I: Consent to Participate

CONSENT TO PARTICIPATE IN NON-BIOMEDICAL RESEARCH

Modified Cooper Harper for Unmanned Vehicle Displays

You are asked to participate in a research study conducted by Professor Mary Cummings Ph.D., from the Aeronautics and Astronautics Department at the Massachusetts Institute of Technology (M.I.T.). You were selected as a possible participant in this study because the population this research will influence is expected to contain men and women between the ages of 18 and 50 with an interest in using computers. You should read the information below, and ask questions about anything you do not understand, before deciding whether or not to participate.

• PARTICIPATION AND WITHDRAWAL

Your participation in this study is completely voluntary and you are free to choose whether to be in it or not. If you choose to be in this study, you may subsequently withdraw from it at any time without penalty or consequences of any kind. The investigator may withdraw you from this research if circumstances arise which warrant doing so.

• PURPOSE OF THE STUDY

The study is designed to evaluate how effectively some surveys help in classifying, ranking and rating unmanned vehicle displays.

• PROCEDURES

If you volunteer to participate in this study, we would ask you to do the following things:

- Attend a training session to learn how to complete the unmanned vehicle mission and review the survey you will be asked to complete after the mission is complete (estimated time 0.25 hours).
- Practice on the unmanned vehicle display until you can complete unmanned vehicle missions (i.e. search for and locate victims or search for and track targets) (estimated time 0.33 hours).
- Execute one 15 minute test trial (estimated time 0.25 hours)
- Attend a debriefing to determine your subjective responses and opinion of the software (estimated time 0.25 hours).
- Testing will take place in MIT building 37, room 301 or MIT building 35, room 220.
- Total time: 1-1.5 hours.

• POTENTIAL RISKS AND DISCOMFORTS

There are no anticipated physical or psychological risks in this study.

- **POTENTIAL BENEFITS**

While there is no immediate foreseeable benefit to you as a participant in this study, your efforts will provide critical insight into the ability of surveys to help in the design of unmanned vehicle displays.

- **PAYMENT FOR PARTICIPATION**

You will be paid \$10/hr to participate in this study. This will be paid upon completion of your debrief. Should you elect to withdraw in the middle of the study, you will be compensated for the hours you spent in the study. In addition, the top performer in this experiment will win a \$50 gift card to Best Buy.

- **CONFIDENTIALITY**

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law. You will be assigned a subject number which will be used on all related documents to include databases, summaries of results, etc. Only one master list of subject names and numbers will exist that will remain only in the custody of Professor Cummings.

- **IDENTIFICATION OF INVESTIGATORS**

If you have any questions or concerns about the research, please feel free to contact the Principal Investigator, Mary L. Cummings, at (617) 252-1512, e-mail, missyc@mit.edu, and her address is 77 Massachusetts Avenue, Room 33-305, Cambridge, MA 02139. The post doctoral investigator is Birsen Donmez at (617) 253-0993, email, bdonmez@mit.edu. The student investigator is Hudson D. Graham at (617) 253-0993, email, hgraham@mit.edu.

- **EMERGENCY CARE AND COMPENSATION FOR INJURY**

If you feel you have suffered an injury, which may include emotional trauma, as a result of participating in this study, please contact the person in charge of the study as soon as possible.

In the event you suffer such an injury, M.I.T. may provide itself, or arrange for the provision of, emergency transport or medical treatment, including emergency treatment and follow-up care, as needed, or reimbursement for such medical services. M.I.T. does not provide any other form of compensation for injury. In any case, neither the offer to provide medical assistance, nor the actual provision of medical services shall be considered an admission of fault or acceptance of liability. Questions regarding this policy may be directed to MIT's Insurance Office, (617) 253-2823. Your insurance carrier may be billed for the cost of emergency transport or medical treatment, if such services are determined not to be directly related to your participation in this study.

• **RIGHTS OF RESEARCH SUBJECTS**

You are not waiving any legal claims, rights or remedies because of your participation in this research study. If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E25-143B, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253 6787.

SIGNATURE OF RESEARCH SUBJECT OR LEGAL REPRESENTATIVE

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

Name of Subject

Name of Legal Representative (if applicable)

Signature of Subject or Legal Representative

Date

SIGNATURE OF INVESTIGATOR

In my judgment the subject is voluntarily and knowingly giving informed consent and possesses the legal capacity to give informed consent to participate in this research study.

Signature of Investigator

Date

Appendix J: Demographic Survey

MCH-UVD DEMOGRAPHIC SURVEY

1. Age: _____

2. Gender: Male Female

3. Occupation: _____

If student:

a. *Class Standing:* Undergraduate Graduate

b. *Major:* _____

If currently or formerly part of any country's armed forces:

a. *Country/State:* _____

b. *Status:* Active Duty Reserve Guard Retired

c. *Service:* Army Navy Air Force Other _____

d. *Rank:* _____

e. *Years of Service:* _____

f. *Specialty:* _____

4. Are you color blind?

Yes

No

If yes:

Which type of color blindness (if known) _____

5. Do you have experience with remotely controlled vehicles (land, sea, air)?

Yes

No

If yes:

a. *Vehicle type(s)/class(es):*

b. *Number of hours:* _____

6. How often do you play video games?

Never

Less than 1 hour per week

Between 1 and 4 hours per week

Between 1 and 2 hours per day

More than 2 hours per day

Appendix K: Evaluation Form for MCH-UVD Rating

MCH-UVD RATING

What MCH-UVD rating would you give this simulator's interface? _____ Why?

Are there any additional problems with the display?

Appendix L: Post Test Debrief

POST TEST DEBRIEF QUESTIONS

- 1. Did the MCH-UVD help you identify deficiencies in the unmanned vehicle display (UVD)? If so, how?

- 2. Did the MCH-UVD help you identify all of the deficiencies in the UVD? If not, what deficiencies were not identified by the MCH-UVD?

- 3. What deficiency did the MCH-UVD help you identify and how important would you say the deficiency was?

- 4. Do you think the 1 to 10 scale of the MCH-UVD is appropriate and effective? Why or why not?

5. In general, how would you change the MCH-UVD to better identify deficiencies in the UVD?

6. What is the best and most helpful aspect of the MCH-UVD?

7. What is the worst and least helpful aspect of the MCH-UVD?

8. Would you have recognized these deficiencies in the UVD without the help of the MCH-UVD? If not, why not?

9. Is the language and appearance of the MCH-UVD clear and understandable? If not, what is not clear?

10. Please express any other comments you may like to share:

Appendix M: MCH-UVD Rating Data

Type of Vehicle: UAV – Condition: Specific

SUBJECT NUMBER	DEMOGRAPHIC DATA			RATING DATA		
	Age	Gender	Occupation	Rating	WHY?	ADDITIONAL PROBLEMS?
1	21	Male	Student	3	Had trouble seeing the duration settings to allow for errors not to occur.	* Layout could be a little more user friendly.
6	19	Male	Student	3	It was hard to expand in the flight path once it was set. It was inefficient that the user couldn't stack commands in an order and have the UAV carry them out 1 by 1.	* Slow.
8	18	Male	Student	5	When you want to change directions, you have to make sure all the settings are correct, and by that time, the vehicle drifts off. You can't precisely control the vehicle.	* You can't queue tasks in order to go straight from one to another.
11	20	Female	Student	3	I just searched randomly because there was no way of seeing if your search was on track before you located a ship.	* Slow.
13	19	Male	Student	5	Not interactive enough – especially with respect to time.	
15	21	Female	Student	4	The display seemed to assume that tracked vehicles would continue in a straight path – but some of them turned. Maybe an approximation of the radius within which the target lies would have been more helpful.	* The tracking tasks should automatically delete. * Also, the tracking vehicle should know not to fly over land – only over the canal.
16	19	Male	Student	8	“Fail to direct operator attention in a timely manner” – ship detection not easily recognizable. Visualization is meaningless and probably reason for slow reaction – bogs down system.	
19	18	Female	Student	4	Even if I had run directly parallel to the ship it will not detect it. The scope of visibility for my ship is not only too limited, but works against me as ships pass me and I continue to search in another direction.	
22	19	Female	Student	6	Random search, most by chance.	* Very simple, which is also good, but otherwise doesn't capture/maintain one's attention.
24	20	Male	Student	3	The trajectory path of the helicopter was very uncertain in its wavering. However, it did get to the target that was set.	

Type of Vehicle: UAV – Condition: Specific

SUBJECT NUMBER	DEMOGRAPHIC DATA			RATING DATA		
	Age,	Gender,	Occupation	Rating	WHY?	ADDITIONAL PROBLEMS?
25	21	Male	Engineer	4	Doesn't tell you how UAV will fly. Direction / speed. Slow to change direction costs valuable search time. Trajectory desired never same as actual trajectory.	* Assigning tasks a pain. Should be able to put several waypoints for allotted fuel. * Slow reaction/performance. * Everything is really spread out.
26	19	Male	Student	5	It did take multiple steps to change search patterns and track targets.	* Mouse was sluggish.
27	19	Male	Student	6	The display and user interface didn't allow quick waypoint additions or changes. Moreover, the predicted route was not anywhere close to the real route. I never could determine the site (width) of the area being searched beneath my UAV.	* Slow. * You have to drag the "Delay penalty" slider instead of clicking a button. * I wish I knew how many seconds of gas I had.
28	20	Male	Student	2	The simulation was adequate for keep track of ships that were located.	
29	18	Female	Student	3	Track patterns were not very precise.	* Slow.

Type of Vehicle: UAV – Condition: General

SUBJECT NUMBER	DEMOGRAPHIC DATA			RATING DATA		ADDITIONAL PROBLEMS?
	Age	Gender	Occupation	Rating	WHY?	
2	24	Male	Engineer	4	The display did not provide me with consequences of my actions. Or tell me how to track ships during the test. Options were available but confusing.	* Display did not show exact course the UAV would take until it started moving.
3	29	Male	Student	4	I think more awareness is needed specially about the range of inspection and routing strategy.	* The energy bar at the bottom of the UAV was confusing.
4	19	Male	Student	8	Although many controls give many levels of control over what's going on, they make things very confusing. Also, the program seems slow, so there is a lag when acting/reacting.	
5	20	Male	Student	2	The only problem was basically the precision of the path that the UAV followed.	
7	23	Female	Student	2	The only issue I found with the display, not counting the inefficient reaction time of the computer is that the program showed where the UAV had previously traveled to. This can affect decision making because it may make a person to be less likely to retrace their steps to search in the same area.	
9	20	Female	Student	4	The display is weak; there could be more tools other than just 'guess-and-cheek' to find ships. The UAV moved too slowly. There should be some measurement of how much fuel is left on display.	
10	20	Male	Student	2	The display was fine. It's just up to the user to use the interface correctly to find the ships.	* Too slow.
12	20	Female	Student	4	The display is insufficient in that the tasks suggested merely mirror the route I had already selected. In fact, the helicopter strayed from selected path at times, therefore wasting time.	* I found the delay time between calculating a task and allowing me to accept the task insufficient.
14	20	Male	Student	5	Not very precise actual waypoint manipulation would help. Also, the ability to adapt to the calculated plan by modifying flight path is needed. Be able to more accurately follow searching/tracking procedures.	* Task names should be clearer instead of just number, i.e. search01 or track01.
17	21	Female	Student	4	It did not help me with decision aiding because I never went back to a spot where I had been already and where the ship may have gone after I left.	

Type of Vehicle: UAV – Condition: General

SUBJECT NUMBER	DEMOGRAPHIC DATA			RATING DATA		ADDITIONAL PROBLEMS?
	Age	Gender	Occupation	Rating	WHY?	
18	33	Male	Student	2	Overall, the display is effective for mission replanning. There are minor modifications that could be made that would aid in maintaining attention and situation awareness with respect to the location of the ships (i.e. radar telling the operator an approx. proximity). These are only personal preferences.	
20	19	Male	Student	5	It takes time between creating a task and the UAV realizing what the controller wants it to do. The UAV also doesn't fly in a straight line.	
21	19	Female	Student	2	Some of the buttons didn't work but it didn't affect my performance.	
23	20	Female	Student	4	It was difficult to change the direction of the UAV until the task was completed, and even then it had to be timed correctly. With the constant changing location of the ships, it would be more efficient to have a simulation that could move in multiple directions in the same task.	
Pilot 2	21	Male	Student	6	I had to implement every step of the process; there was poor automatic waypointing. It also did a poor job of accepting a prioritizing task, in fact, it crashed.	

Type of Vehicle: UGV – Condition: Specific

SUBJECT NUMBER	DEMOGRAPHIC DATA			RATING DATA		
	Age	Gender	Occupation	Rating	WHY?	ADDITIONAL PROBLEMS?
1	20	Male	Student	5	Originally, I considered giving the simulator a 4. However, the words “timely manner” showed through when I was reviewing 5. The time delay was a major factor in the ability to locate the victims quickly.	* The HUD may have been more useful with a real-time radar system. I realize that the purpose of the maps given was to help locate barriers such as walls and desks, but with a faster refresh of relevant environment data, navigation ability would improve.
3	20	Male	Student	4	It would help if the UGV had more than one camera to same time getting a picture of your surroundings, without having to rotate as much. The fact that the map is laid out by the UGV based on noise sensors prevents the operator from successfully predicting what exactly lies ahead at a distant place.	* The map Data Viewed becomes a bit cluttered with all the paths and previous trajectories stored on it.
4	18	Female	Student	4	The display is confusing because turning the teleoperate to look in a different direction doesn’t change the orientation of the robot, making it hard to decide what direction to go in.	* No sense of scale, size of the robot is, how far away things are.
6	20	Male	Student	2	It sometimes seemed like re-planning my search on the spot was difficult, such as when I wanted to turn a vehicle around after bumping into a wall, but that was just due to my lack of experience. With more practice, I would have got it better.	
7	19	Male	Student	4	Map Data Viewer: Difficult to locate (depth perception) on map when compared to video feed. Victims appear closer in video feed than on map.	* Unreasonable to suggest military wouldn’t acquire detailed plans of the building before evacuation. Condition may be unknown, but I believe general layout would oftentimes be known.
9	20	Male	Student	3	The fact that you could change the camera angle made it difficult to decipher whether (on the map) the robot was oriented toward the camera or the wheels.	* No, just the delay in the technology and the slowness of the robot made the mission difficult.
10	25	Female	Student	5	The added step of clearing the paths unnecessary. After selecting the UGV you should be able to directly choose the path.	* I don’t see why there is two displays. One could make it easier, or an option to turn one off.
11	20	Female	Student	4	Although borders on the map were clear, entrances to rooms, pathways, etc did not make themselves clear (appear grey) until the UGV passed said location, thus forcing me to change paths midway. Although this would be true in a real life situation, it made navigation a bit hard at times during the simulation.	

Type of Vehicle: UGV – Condition: Specific

SUBJECT NUMBER	DEMOGRAPHIC DATA			RATING DATA		
	Age	Gender	Occupation	Rating	WHY?	ADDITIONAL PROBLEMS?
12	19	Female	Student	4	Able to locate victims with simulator. The display shows victims to be located, however, is not as effective as it could be. The display does not show objects like desks or other obstructions.	* Does not show enough detail which makes it more difficult to maneuver to discover victims.
13	19	Female	Student	5	I was able to locate victims, but only after readjusting location and view several times for each one which took a lot of time.	* Time lag, which made it confusing as to which way was which.
15	20	Male	Student	6	Sometimes I had to do several different missions to get to the desired location. I had to make a lot of estimate of where victims were. Map is not clear.	* It is very hard to tell the robot where the camera should point in a mission. * Map is not clear in showing doorways.
22	18	Male	Student	5	To find victims, I had to estimate where they were on the map to find them, unless they were in the open. The mapping was minimal and people were shaded the same as walls. The robots were sluggish in response and did not always follow waypoints exactly.	
25	20	Female	Student	3	The victims were easy to locate with help from the visual, but being exact was hard. All the screens/windows worked well together.	* The MAPDATA Viewer got a little crowded when I was trying to place the victims because it showed the vehicle movement.
26	19	Male	Student	3	Although the display was quite straightforward, the limited angle of camera panning and slow speed of camera movement hindered performance.	* Other than those mentioned above, all problems were with UGC intelligence (difficulty navigating doorways, even when given sufficient waypoints).
29	21	Female	Student	3	The time delay is confusing and the robots movement was slow. The camera angles were also slow to change. It was hard to locate victims.	* I found four robots display confusing. Switching between displays was also confusing.

Type of Vehicle: UGV – Condition: General

SUBJECT NUMBER	DEMOGRAPHIC DATA			RATING DATA		ADDITIONAL PROBLEMS?
	Age, Gender, Occupation	Rating	WHY?			
2	21	Male	Student	3	The decision to go manual once there were problems with going automatic caused some problems. Also, the manual controls were not that great. Constantly used automatic drive mode.	
5	19	Male	Student	3	All decisions were fairly simple to implement, but it could be tedious to customize it to my need.	* Minor one, a new feature actually. Instead of rotating one camera on vehicle, multiple cameras could be used to capture videos from all directions (to save time from rotating).
8	21	Female	Student	5	Automatic pop-ups were distracting when planning missions with other vehicles. Allowing multiple/saved viewpoints per mission would be more useful for accessing grey areas and marking people.	
14	19	Male	Student	8	The display should be easier to see and the UVs should move more smoothly. It however accurate makes the whole thing slightly more confusing.	
16	18	Male	Student	4	Main video display was not very useful, the smaller ones were enough. The blue background was bright and distracting. Using dark grey might be better. Also having to clear robot's orders before issuing new ones even though they were completed was annoying. They should clear automatically.	* Being able to select the robots on the top map would be convenient. * Also labeling victims on top map would be easier. * Really only one map is necessary.
17	21	Female	Student	3	The task was not hard but it took long to respond to my commands.	
18	21	Female	Student	5	The display reacts slowly which in turn causes the operator's reactions to be slow. This hinders decision making and rapid changes which need to be made.	* No. If the speed was quicker I think the display is efficient.
19	21	Male	Student	2	Display is a bit slow when changing direction of camera. The icon for changing the direction of the camera was also annoying to use.	
20	20	Male	Student	5	Effective but extremely inefficient, needs improvement in most areas. Basic information is available.	* Slooooooow. * Lag. * Low resolution. * Map rotates w/ robot. * Need to be able to rotate map.

Type of Vehicle: UGV – Condition: General

SUBJECT NUMBER	DEMOGRAPHIC DATA Age, Gender, Occupation			RATING DATA		
				Rating	WHY?	ADDITIONAL PROBLEMS?
21	20	Male	Student	4	Because I agree w/ the paragraph.	<ul style="list-style-type: none"> * It doesn't tell you when you're hitting something. * It doesn't tell you when a UV is idle. * By the end it got cluttered. * It was easy to learn but still not so easy to operate. * It should alert / not allow when two vehicles will collide. * It should automatically fill in gaps in map. * Buildings have blueprints – these things should already have maps.
23	21	Female	Student	3	Display is easy enough to use, camera seems slower to respond than the vehicle movement responds.	<ul style="list-style-type: none"> * Directing the camera after setting waypoints was the hardest task I found.
24	21	Female	Student	3	Display allows for accurate judgment of distance from objects + victims and also allows easy identification of all possible areas to be searched.	<ul style="list-style-type: none"> * Time delay.
27	45	Male	Engineer	6	Needed to determine relationships of robots. – If the robots were physically colored to match the display it would increase situational awareness. – The question was (for me) I can see another robot – Which one is it?	
28	18	Male	Student	4	It was pretty simple to use but some symbols like the triangles seem unnecessary. It also doesn't let you know when something is blocking the way.	<ul style="list-style-type: none"> * None besides something to better define obstacles and to reduce the lag if possible.
30	20	Female	Student	2	One of my robots got stuck. I couldn't figure out how to get it out. Otherwise, simulator was easy to use.	<ul style="list-style-type: none"> * Would be easier if the graphics moved faster – seemed like there was a lot of time spent waiting for the robots to move.

Appendix N: Post Test Debrief Responses

Type of Vehicle: UAV – Condition: Specific

SUBJECT NUMBER	RESPONSE				
	Did MCH-UVD help identify deficiencies? If so, how?	What deficiencies were not identified by MCH-UVD?	Best (most helpful) aspect & other positive comments	Worst (least helpful) aspect & other negative comments	Would recognize deficiencies without MCH-UVD?
1	Yes.	Not general enough. e.g., the lack of weather information.	Best – Flowchart. *Language and appearance good.	Worst – No rating in between the suggested levels.	Yes. But cannot assign importance.
6	Yes. Provided criteria – easy for a person who does not know what to look for.	It did not identify some of the deficiencies.	Best – Flowchart. *Language clear, likes the color coordination.	Could have more numbers, expand flowchart to include more criteria.	No. Only some.
8	Yes. Helpful in making suggestions.	It just helped identify the particular number.	Best – Descriptions. *Language and appearance good.	Worst – Flowchart questions.	Yes. But less specific.
11	Yes. Narrow down specific problems.	Display too slow.	Best – Descriptions. *Language and appearance good.	None.	Yes. But not specific.
13	Yes. Qualify initial concerns.	Display slow, not responsive.	Language and appearance good.	Worst – Flowchart questions. Just read from the bottom.	No. Only some.
15	Yes. Good descriptions.	Had to pick one number.	Best – Color scheme.	Worst – Terms. *Underline keywords.	Yes. But sheet was helpful.
16	Yes. Gave him way to quantify and break down into groups.	None.	Best – Descriptions. *Language and appearance good.	Worst – Color.	Yes. But not with the same terms.
19	Yes. Flowchart – but descriptions are too specific.	A few.	Best – Flowchart. *Language and appearance good.	Worst – Descriptions too specific, not enough variety.	Yes. But it gave good direction.
22	Yes. Didn't have idea until read it.	A few.	Best – Second column. *Language and appearance good.	Worst – Questions are too simple.	Yes. But cannot describe it as well.
24	Yes. Didn't pay attention to deficiencies but sheet made him realize.	The minor ones.	Best – Flowchart. *Language and appearance good.	Worst – Categories.	No.
25	Yes. But many things in different categories that he liked.	More than what you would have seen initially.	Best – Descriptions. *Language and appearance is good.	Worst – Flowchart questions.	No. Only most of them.

Type of Vehicle: UAV – Condition: Specific

SUBJECT NUMBER	RESPONSE				
	Did MCH-UVD help identify deficiencies? If so, how?	What deficiencies were not identified by MCH-UVD?	Best (most helpful) aspect & other positive comments	Worst (least helpful) aspect & other negative comments	Would recognize deficiencies without MCH-UVD?
26	Yes. Described what he thought was wrong.	Identified deficiencies that were part of the display not just computer.	*Language and appearance good.	Worst – Wording too difficult.	Yes. But no rating.
27	Yes. Easy to follow questions (definite answer), terms.	Display slow.	Best – Flowchart questions. *Language and appearance is good.	Worst – Descriptions too long.	Yes. But couldn't have categorized them or indicate severity.
28	No.	NA	Best – Directs reasoning. *Language and appearance is good.	Worst – Flowchart was unnecessary. *Need more specific categories.	Yes.
29	Yes. Wrote out problems that could be wrong.	None.	Best – Terms. *Language and appearance good.	Worst – Flowchart questions.	Could have done better without it.

Type of Vehicle: UAV – Condition: General

SUBJECT NUMBER	RESPONSE				
	Did MCH-UVD help identify deficiencies? If so, how?	What deficiencies were not identified by MCH-UVD?	Best (most helpful) aspect & other positive comments	Worst (least helpful) aspect & other negative comments	Would recognize deficiencies without MCH-UVD?
2	Yes.	None.	Best – Flowchart. *Language and appearance good.	Worst – Lots of reading.	Yes. But cannot pinpoint to one thing.
3	Yes. Categorized different levels.	Deficiencies related to the path planning algorithm, and the imprecise flight path	Best – Categorization of difficulties. *Language and appearance good.	None.	Yes.
4	Yes. Breaks down what could be wrong in the display. Helps you think through.	More things in other numbers.	Best – Reading through all descriptions uncovers all deficiencies. *Language and appearance good.	Worst – Questions in flowchart, got to top through questions, but descriptions at the bottom were more accurate *should have less numbers, and more broad categories.	Yes. But not so specifically.
5	Yes.	None.	Best – Terms.	Worst – Some wording. *Language not too clear.	Yes. Because it was a minor deficiency.
7	No.	NA	Best – Color coordination and terms. *Appearance is good.	Worst – Run-on sentences, wording is bad.	Yes.
9	Yes.	Pace.	Best – Putting it into words *Language clear, appearance is good.	Worst – Too easy to get high rating.	Yes. Just not same words.
10	Yes.	None.	Best – Terms. *Language and appearance good.	Worst – Flowchart too broad, descriptions too wordy.	Yes. Because there were not many.
12	Yes. Knew what problems were and just looked for correct description – got to what she was thinking.	Layout of interface.	Best – Flowchart if correctly categorized.	Worst – Colored boxes. *Language: questions are poorly worded.	Yes. But cannot assign a rating.
14	Yes. Problems in multiple categories.	Had to pick one number.	Best – Descriptions. *Language and appearance good.	Worst – Diagnosis terms.	Yes. But no rating.

Type of Vehicle: UAV – Condition: General

SUBJECT NUMBER	RESPONSE				
	Did MCH-UVD help identify deficiencies? If so, how?	What deficiencies were not identified by MCH-UVD?	Best (most helpful) aspect & other positive comments	Worst (least helpful) aspect & other negative comments	Would recognize deficiencies without MCH-UVD?
17	No.	NA	None. *Language and appearance good.	*Suggest more categories that are more specific. Need examples.	Yes.
18	Yes. Guideline to say what affected his performance, leads you to say specific things.	Some of them.	Best – Flowchart. *Language and appearance good.	None.	No. Was very helpful.
20	No. Already knew them.	NA	Best – Flowchart, prioritizes deficiencies successfully, colors are good. *Language and appearance good.	Worst – Descriptions.	Yes.
21	No.	NA	None.	Worst – Flowchart. Should have been lower rating.	Yes.
23	Yes. But limiting because it could be in between.	Helped identify major ones but could be other ones in lower categories.	Best – Organization. *Language and appearance good.	Worst – Similarities in descriptions, too close. *More categories more variations needed. Broader but more detailed.	No.
Pilot 2	Yes. Ranked which issues should be fixed first.	None.	Best - Diagnosis of display issues (terms) * Language very clear.	Worst – Rating scale	Yes.

Type of Vehicle: UGV – Condition: Specific

SUBJECT NUMBER	RESPONSE				
	Did MCH-UVD help identify deficiencies? If so, how?	What deficiencies were not identified by MCH-UVD?	Best (most helpful) aspect & other positive comments	Worst (least helpful) aspect & other negative comments	Would recognize deficiencies without MCH-UVD?
1	Yes. The flowchart makes it simple and easy to navigate through the different ratings.	Does not seem to go over the manual procedure of actually doing the tasks and operating the vehicles.	Best – The separation through color coding and grouping. *Appearance good.	Worst – Some ratings are too general (e.g., 5). *More rankings might be appropriate to include more specific evaluations. *Language needs to be simplified for people who are not familiar with the terms.	Yes. But probably not as accurately.
3	Yes. It lists them into a set of categories that effectively point out the areas that could be improved.	Functionalities of the UGV.	Best –The flowchart. Language and appearance good.	Worst – Some categories are too condensed. *The ratings can be broken up into more levels. Also, add other ratings to deal with UGV functionalities.	No. Because he has not used the display long enough.
4	Yes. It categorized the problems – easier to identify where the problems are.	None.	Best – The flowchart.	Worst – The numbered ratings. *More ratings needed. *The language is too technical.	Not all of them.
6	Yes.	None.	Best – The flowchart. *The color coding.	Worst – The wording. *More ratings needed. *Language not clear, sentences are long.	No. It would require a lot of time.
7	Yes. It lists all the possible problems.	A few.	Best – The flowchart. *Appearance is good.	Worst – The wording (too technical) and the broadness of each rating. It should be more specific with shorter sentences. *Do a list format rather than paragraph. *For rating 4, there should be more categories.	Yes.
9	No.	The procedure to mark victims. The small thumbnail pictures for each UGV.	Best – The flowchart.	Worst – The petition of words which makes some of the ratings too similar to pick from. *It should be a smaller scale. The differences between some ratings are too minute (i.e., “7 and 8”, “3 and 4”). *There should be different categories, more questions that ask things like “how helpful each aspect of the interface way”, then a tally up answer and rate depending on final score. *Too wordy.	Yes.

Type of Vehicle: UGV – Condition: Specific

SUBJECT NUMBER	RESPONSE				
	Did MCH-UVD help identify deficiencies? If so, how?	What deficiencies were not identified by MCH-UVD?	Best (most helpful) aspect & other positive comments	Worst (least helpful) aspect & other negative comments	Would recognize deficiencies without MCH-UVD?
10	Yes. It lists the possible deficiencies, and helps categorize.	More options to customize.	Best – The severity rating.	Worst – The language used. But if you are used to the terminology it would be fine. *Need more ratings, and details. *If you notice something else, there should be room to add in to the evaluation.	Yes. But it helped her categorize.
11	Yes.	None.	Best – Flowchart. *It prioritized the problems, well organized. *Appearance is good.	Worst – Some of the descriptions were confusing or were hard to differentiate, so it took longer to decide what rating was right. *Language was confusing.	No.
12	Yes. It listed the possible problems.	None.	Best – Flowchart format. *Appearance is good.	Worst – The second column. *The flowchart questions are too general. It is not always “yes” or “no”. *Language.	Yes.
13	Yes. It presented problems she hadn’t thought of.	None.	Best – Flowchart. Makes it faster. *Language and appearance is good.	Worst – Third column. *Should be able to select multiple ratings. Make it a checklist.	No. Also, couldn’t have been able to explain them.
15	No.	Positioning of the robots at the end of each task. Robots running into each other. Clutter.	Best – The column with the descriptions.	Worst – The rankings are not accurate (e.g. 5 is not necessarily better than 7). *Flowchart leads you to wrong place. Eliminate it and have the person read all the ratings, and decide which one applies the most. *Language not clear, with long sentences.	Yes.
22	Yes. It helped him put the problems into words and designate how important they were.	The issues with the map.	Best – The fact that it led you in the general direction. *Appearance is good.	Worst - Selecting only one rating. Flowchart led him to the wrong rating. Use a checklist. *Language too wordy.	Yes.
25	Yes. It listed the things you should be able to do with the UVD and you could point out what it was lacking.	Moving the UGVs around the walls. It was mostly focused on finding the victims.	Best – The descriptions. *Language and appearance is good.	Worst –Third column. *Have more branches, more options.	No. She wouldn’t know what to look for.

Type of Vehicle: UGV – Condition: Specific

SUBJECT NUMBER	RESPONSE				
	Did MCH-UVD help identify deficiencies? If so, how?	What deficiencies were not identified by MCH-UVD?	Best (most helpful) aspect & other positive comments	Worst (least helpful) aspect & other negative comments	Would recognize deficiencies without MCH-UVD?
26	Yes. It lists the problems and the strengths.	None.	Best – The flowchart questions to help you find what rating you are looking for. *Language and appearance is good.	Worst – Flowchart questions can be too general at times.	Yes.
29	Yes. It suggested the stuff that could be wrong with it.	None.	Best – The flowchart questions.	Worst – The descriptions. Not clear. Sentence structure is confusing. Too wordy. *Have a checklist rather than one rating.	Yes.

Type of Vehicle: UGV – Condition: General

SUBJECT NUMBER	RESPONSE				
	Did MCH-UVD help identify deficiencies? If so, how?	What deficiencies were not identified by MCH-UVD?	Best (most helpful) aspect & other positive comments	Worst (least helpful) aspect & other negative comments	Would recognize deficiencies without MCH-UVD?
2	Yes. It allows differentiating between manual and automated.	None.	Best – The flowchart.	Worst – Too technical wording. *The rating scale because some rankings are skipped, may lead to skipping deficiencies.	Yes.
5	Yes. It breaks down the different problems, so it makes it easier to identify the difficulty encountered.	None.	Best – The flowchart.	Worst – The wording. *Language is not clear. Technical terms.	No. It is hard to put the deficiencies into words, and categorize them.
8	Yes. It reminded you of the cognitive processes, maneuverability, and time management.	User options. Selecting where the UGV points at the end of the mission.	Best – Flowchart. *Not used to the language but she was able to figure out. *Appearance is good.	Worst – The second column. *The severity of ratings is not accurate. Some seem equal but different. No need for ranking. *Add more to the flowchart to get your mind flowing. *A space is needed to write in specific examples.	No. It would have been hard to break down the thought process.
14	No.	Lag. Slow movement of robots. Confusing maps.	Best – The descriptions helps you think of problems. *Appearance is good.	Worst – Confusing questions with no straightforward answers. *Some of the questions are too vague. There should be a more than “yes” or “no”, maybe a third path you can take. *Language not clear (e.g., safely completed).	Yes.
16	Yes. It helped him think about the problems, but also the strengths of the display.	Video feedback.	Best – The definitions, and color coding. *Appearance is good.	Worst – The second column. *Maybe have a scale for each display feature (e.g., video). *Language is not clear.	Yes. But cannot assign severity.
17	No.	Slow response.	Best – The visual presentation. *Appearance is good.	Worst – Too wordy. *Some ratings were not clear, and some very similar. Have more general ratings to include more problems.	Yes.
18	Yes. Things that could have occurred were listed.	None.	Best – The definition of the problems. *Language and appearance is good.	Worst – The second column. *Read all of the ratings.	No.

Type of Vehicle: UGV – Condition: General

SUBJECT NUMBER	Did MCH-UVD help identify deficiencies? If so, how?	What deficiencies were not identified by MCH-UVD?	RESPONSE		
			Best (most helpful) aspect & other positive comments	Worst (least helpful) aspect & other negative comments	Would recognize deficiencies without MCH-UVD?
19	Yes. It pinpointed the problems.	The camera rotation control.	Best – The descriptions. *Appearance is good.	Worst – The flowchart questions are not too clear sometimes. There is no alternate answer between “yes” or “no”. *Read all the ratings. *Too technical.	Yes.
20	Yes. It asked questions about details I haven’t thought of before.	Slow response.	Best – The descriptions. *Language and appearance is good.	Worst – The flowchart questions are vague, did not lead to the correct branches. *He changed his rating after reading all ratings.	No. He does not have enough experience with the display.
21	Yes. It led him to the problems.	None.	Best – It makes you make a decision.	Worst – The flowchart led me to the wrong rating. Don’t choose only one rating. *Allow the operator to add in comments. *Language is sometimes ambiguous. Too technical.	No. Not all of them.
23	Yes.	Speed of the camera.	Best – The flowchart questions.	Worst – The wording should be in the first person. It is understandable but too wordy. *She would like 10 to indicate the best display. *The scale should be on a bigger sheet.	No. It would take longer.
24	Yes. By reading the demands on the operator, you realize how the display affected your decision.	Slow response.	Best – It outlines the demands required of you that you might not be able to articulate yourself. *Language and appearance is good. *Like the color-coding.	Worst – Some of the ratings seem to similar and can be put together (e.g., 5 and 6). *Give examples along with the descriptions.	Yes.
27	Yes. The flowchart led him in the right direction, and the descriptions were clear and they applied.		Best – Descriptions. *Language and appearance is good.	Worst – Change the flowchart questions to have another layer in addition to “yes” and “no”. *Reverse the ratings to have 10 be the best.	Yes.

Type of Vehicle: UGV – Condition: General

SUBJECT NUMBER	Did MCH-UVD help identify deficiencies? If so, how?	What deficiencies were not identified by MCH-UVD?	RESPONSE		Would recognize deficiencies without MCH-UVD?
			Best (most helpful) aspect & other positive comments	Worst (least helpful) aspect & other negative comments	
28	Yes.	None.	Best – The grouping of the ratings. Flowchart. *Language and appearance is good.	Worst – Reverse the ratings to have 10 be the best.	Yes.
30	No.	Slow response. Robots getting stuck.	Best – The flowchart questions. *Language is good.	Worst – Descriptions are too vague.	Yes.