AIAA AVIATION Forum
August 2-6, 2021, VIRTUAL EVENT
AIAA AVIATION 2021 FORUM

10.2514/6.2021-2330

# A Cognitive Walkthrough of Multiple Drone Delivery Operations

Casey L. Smith*, Garrett G. Sadler†, Terence L. Tyson*, Summer L. Brandt*, R. Conrad Rorie†, Jillian N. Keeler‡
*NASA Ames Research Center, Moffett Field, CA, 94035, USA*

Kevin J. Monk§
*NASA Armstrong Flight Research Center, Edwards, CA, 93523, USA*

Igor Dolgov**
*Joby Aviation, Santa Cruz, CA, 94070, USA*

Jesus Viramontes††
*California State University, Long Beach, CA, 90840, USA*



**Advances of early twenty-first century aviation and transportation technologies provide opportunities for enhanced aerial projects, and the overall integration of unmanned aircraft systems (UAS) into the National Airspace System (NAS) has applications across a wide range of operations. Through these, remote operators have learned to manage several UAS at the same time in a variety of operational environments. The present work details a component piece of an ongoing body of research into multi-UAS operations. Beginning in early 2020, NASA has collaborated with Uber Technologies to design and develop concepts of operations, roles and responsibilities, and ground control station (GCS) concepts to enable food delivery operations via multiple, small UAS (sUAS). A cognitive walkthrough was chosen as the method for data collection. This allowed information to be gathered from UAS subject matter experts (SMEs) that could further mature designs for future human-in-the-loop (HITL) simulations; in addition, it allowed information to be collected remotely during the stringent restrictions of the COVID-19 pandemic. Consequently, the described cognitive walkthrough activity utilized remote data collection protocols mediated through the usage of programs designed for presentation and telecommunications. Scenarios were designed, complete with airspace, contingencies, and remedial actions, to be presented to the SMEs. Information was collected using a combination of rating scales and open-ended questions. Results received from the SMEs revealed expected hazards, workloads, and information concerns inherent in the contingency scenarios. SMEs also provided insight into the design of GCS tools and displays as well as the duties and relationships of human operators (i.e., monitors) and automation (i.e., informers and flight managers). Implications of these findings are discussed.**

* Research AST, Human Systems Integration, MS262-2.
† Research AST, Human Systems Integration, MS262-2, AIAA Member.
‡ Research Student Trainee, Human Systems Integration, MS262-2.
§ General Engineer, Human Systems Integration. MS262-2.
** Human Factors Engineer, Air-Taxi Products, 340 Woodpecker Ridge.
†† Graduate Student, Department of Psychology, 1250 Bellflower Boulevard.

# I. Introduction

Since their initial appearance as "aerial torpedoes" and "radio-controlled weapons delivery platforms" in the early 1900s, unmanned aircraft systems (UAS) have been advanced and transformed by aeronautical projects undertaken and technologies developed worldwide [1]. By the early 2000s, unmanned aerial vehicles (UAV) were utilized for recreation and special interest activities as well as a range of government and industry operations like power/gas line inspection, surveying, agriculture, surveillance, law enforcement, and emergency search and rescue [2]. With such a diverse body of applications, it should not come as a surprise that UAS operations feature a corresponding diversity in the potential arrangements of vehicles, professionals, and advanced software and automation. As such, this has led to commercial and military applications in which multiple UAS are managed by a single operator. Recent projects have explored ground control station (GCS) configurations as well as human autonomy teaming (HAT) principles that more efficiently allow oversight of multiple (N) aerial vehicles by multiple (M) operators, known as M:N operations [3]. Further research is required to study such novel configurations of vehicles and personnel—i.e., to study Concepts of Operations (ConOps) and the Roles and Responsibilities (R&R) of the concomitant aviation professionals—as the operational environment moves from the traditional air traffic management (ATM) domain towards newer concepts like UAS Traffic Management [2].

## A. Unmanned Aircraft Systems (UAS) Traffic Management

The Federal Aviation Administration (FAA) defines the small unmanned aircraft systems (sUAS) category as all UAS weighting more than 0.55 pounds and less than 55 pounds. Using aircraft registration data and industry sales figures, the FAA estimated in 2018 that there were approximately 1.5 million sUAS operating in the United States [4], composed of approximately 1.25 million model/recreational sUAS and 277,000 non-model/commercial sUAS. The FAA expects that this combined fleet of recreational and commercial sUAS will grow to a size of between 2 to 3 million by 2023, with growth concentrated in the commercial sUAS markets. At that time, the FAA expects the fleet size of commercial operations to be three times the size of non-commercial operations [4]. This dramatic increase to the volume of domestic UAS operations is expected to yield a consequential and corresponding demand for new, dedicated airspace services to support them.

Anticipating the need to develop a means for safely integrating novel types of sUAS operations into controlled and uncontrolled areas of the United States' National Airspace System (NAS), NASA began work on the UTM conceptual framework in 2013 [2]. The UTM ConOps was designed as a federated system to support the projected increase of sUAS flights at low altitudes (i.e., below 400 feet above ground level [AGL]). The system is made up of UAS operators, a network of UAS Service Suppliers (USSs), a Flight Information Management System (FIMS), and Supplemental Data Service Providers (SDSPs), with regulatory oversight by the FAA. Within this framework, the UAS operator is the entity who oversees the management of their operation: they must meet regulatory responsibilities, plan flights, share operation and intent information, and safely conduct flights using all available information from the greater UTM system. Each UAS operator makes use of a USS, which assists in meeting operational requirements by acting as a communication bridge between the federated operators and the broader UTM system. The services provided by a USS support planning, intent sharing, strategic deconfliction, conformance monitoring, remote identification, airspace authorization, and airspace management functions. FIMS provides an interface for data exchange between the overall FAA data management systems and the UTM ecosystem, and UAS operators can optionally subscribe to SDSPs to receive additional information such as enhanced data services for terrain, obstacles, and specialized weather. The person ultimately responsible for the safe conduct of UAS flight(s) is known as the remote pilot in command (RPIC). Unlike conventional Air Traffic Management, UTM operations are not managed by air traffic control (ATC). Instead, the network of linked USSs is intended to provide cooperative management of low-altitude operations without direct FAA handling and assistance.

Subject matter experts (SMEs) within the UTM system obtain a Performance Authorization from the FAA allowing them to conduct flights within the geographic bounds of an Authorized Area of Operations. When a UAS operator wishes to fly, an operation plan will be submitted to the overall USS network via the operator's contracted USS, in the form of a four-dimensional (4D) volume of airspace, mapping out the area in which the flight will operate in both space and time. The network of USSs shares and coordinates submitted operation plans to provide strategic deconfliction services as well as checks for constraint information, passes along advisories (e.g., Notice to Airmen [NOTAMs], traffic conflicts, storms, unexpected obstacles such as temporary structures), and informs operators of special restrictions. During flight, UTM operators are responsible for maintaining separation from: aircraft, restricted areas of airspace (i.e., UAS Volume Reservations [UVRs]), unsafe weather cells, terrain, hazards, and any other unsafe conditions. Operators are required to remain in conformance with their operation plan and for all in-flight coordination with other UAS operators. If, at any point, a flight is outside of conformance or has an on-board equipment problem (e.g., lost link, overheating battery), the operator must correct the problem and return the operation to conformance. In such events wherein the situation is not correctable, the operator must notify affected airspace SMEs as soon as practical and execute a predictable, appropriate contingency plan.

## B. M:N Paradigm

The UTM concept laid out above provides an architecture for enabling sUAS operations; however, UTM itself is generally agnostic to how a UAS operator chooses the assignment, management, and supervision of RPICs and the vehicles they fly. Given the expected growth in demand for large-scale, dense, commercial sUAS operations, the current 2:1 and 1:1 operator-to-vehicle crew configuration traditional to aviation is a likely untenable bottleneck [5]. The burden and cost associated with completing the training and professional requirements for piloting aircraft has motivated a force-multiplication approach wherein a single pilot is responsible for the simultaneous management and/or control of numerous aircraft. Since the mid-2000s, numerous studies have examined the effectiveness of single-pilot control of multiple vehicles [3, 6-9]. Within the human factors literature, researchers have found that one of the most critical factors for achieving effective supervisory control of multiple UAS is the maintenance of adequate situation awareness (SA) of the overall tasking environment as well as SA of each individual vehicle [10]. Additionally, cognitive research shows that task switching—a frequent activity in multiple vehicle control—carries a cost: People's responses tend to be substantially slower and more error-prone after switching between two or more individual tasks [11-12], though there is evidence that this cost may be reduced when SMEs are prepared for the switch or received task-switching cues [13]. Finally, as the number of vehicles increases, and thus the cognitive load on the pilot, mission efficiency can suffer. Porat and colleagues [14] found that simultaneous control of two to four vehicles resulted in pilot difficulty processing information from multiple, separate sources. Similarly, Monk and colleagues [3] observed a greater number of missions completed, but this result came at the expense of individual mission efficiency due to the disparate attention allocated to the assigned missions.

One solution to SA, workload, and task-switching limitations to multiple vehicle control can be found in a natural extension of the single operator, multiple vehicle control paradigm laid out above: the M:N control paradigm. In this configuration, multiple operators share multiple assets between them. Utilizing the M:N crew configuration allows for force multiplication because of its flexibility. Having multiple operators and the ability to pass vehicles between them enables this crew configuration to dynamically adjust to changes/spikes in cognitive workload and to handoff vehicles to specialists (e.g., maintainers, harbor pilots) as the situation requires [5]. Accordingly, the M:N crew configuration provides for a UTM application and ConOps that alleviates some of the challenges found by researchers studying single-operator control of multiple vehicles.

## C. Proposed Roles and Responsibilities for a M:N Crew Configuration

The M:N paradigm is a broad category of control configurations that may be deployed across a variety of operational contexts including UTM, Urban Air Mobility, large UAS operations, high altitude platform systems, and UAS swarms. The concept is meant to enable a future state of scalable operations for increasingly autonomous vehicles. In this study, we employed a type of M:N configuration that posits the personnel role of remote pilot/operator. The remote pilot is responsible for monitoring the airspace, anticipating any encounters with traffic or hazards, anticipating the need to run special contingency operations, and generally attending to associated assets. In an M:N configuration, there is the possibility of M-many operators attending to a total of N-many aircraft between them.

## D. Study Design

A cognitive walkthrough is a process of inspecting the useability of interactive systems to ascertain how well they can be utilized by operators with varied experience and expertise. Cognitive walkthroughs are often accomplished by leading subjects through the tasks involved in a system while simultaneously extracting their feedback about these processes [15]. This cognitive walkthrough method provided the added benefit of allowing us to collect data while also observing strict COVID-19 restrictions.

The SMEs were presented with a series of sUAS urban food delivery scenarios, where they were asked to adopt the perspective of an operator under a 1:9 control configuration. The scenarios differed in terms of the type, and complexity, of contingencies that could unfold over the course of a mission. The goal of the cognitive walkthrough was to elicit feedback from SMEs regarding the general 1:N ConOps, the responsibilities associated with the operator and assistive automation, and the GCS displays and tools. SMEs' responses to the questionnaire scales regarding each use case were analyzed, and transcripts of their interview sessions were processed using constructivist grounded theory protocol (i.e., open coding, focused coding, and theory formation) [16]. Further elaboration on grounded theory as a method, as well as its role in the current study, is beyond the scope of this report, and it is therefore a topic of exploration for future publications. By making these adaptations, feedback from this study was intended to be used to inform future in-person simulations, which would build up from a 1:N control paradigm to an M:N control paradigm.

## II. Method

### A. Participants

Ten SMEs with experience in manned and/or unmanned flight operations were recruited for the virtual cognitive walkthrough ($M_{age}$ = 43 years old, $SD$ = 10.76). Their manned flight time totaled over 14,400 hours in civilian aircraft and over 14,500 hours in military flights; from the military operations, over 4,900 hours were flown in combat. Their unmanned flight time totaled over 4,600 hours in civilian aircraft and over 8,100 hours in military flights; from the military operations, over 7,000 hours were flown in combat.

### B. Environment

Each SME was led through a PowerPoint presentation demonstrating four scenarios that provided examples of how an operator would simultaneously manage multiple sUAS with a proposed GCS configuration. Designs for scenarios, use cases, operator displays, and controls were produced through a series of iterative, collaborative brainstorming sessions between NASA and researchers from Uber Elevate from February to May 2020. Both interface designs and use case scenarios were constructed to avoid overcomplexity; this was so that they would be easier to explain, understand, and inspected by the SMEs during the remote cognitive walkthrough. To further facilitate remote presentation and data collection, video teleconferencing was used through the Microsoft Teams business communication software platform.

#### *1. Operational Concept*

SMEs were directed to envision the task of monitoring nine sUAS in San Diego, California under the UTM framework. The task was described as an investigation into the ConOps developed by the researchers for food delivery with sUAS. The aircraft were said to be equipped with modern capabilities like autopilot, electric vertical takeoff and landing (eVTOL), detect and avoid, and belly mounted cargo transport. The ConOps assumed that aircraft would be designed to fly preprogrammed routes managed by a USS, starting from takeoff and ending when they returned to their base of operations, known as a 'hive.' Each flight lasted approximately 20 minutes, flew beyond visual line of sight (BVLOS), and stayed below 400' AGL. The mission assumed operations involving one hive, one restaurant, and nine delivery locations.

#### *2. Ground Control Station*

The GCS, shown in Fig. 1, was split between two displays. The Timeline and Mission display, shortened as the "Timeline," is positioned on the left. The top of this display provides a temporal representation of the active aircraft, showing their respective mission progress and available battery power. The bottom right side of the Timeline display shows the beginning and end timestamps of the mission as well as the aircraft's system health and conformance status. The bottom left of the screen is reserved for displaying off-nominal events. The Tactical Situation Display (TSD) is the central screen, displaying the aircraft position over a moving map of the airspace in San Diego County; it includes the routes, hives, restaurants, delivery destinations, and any other points of interest. Aircraft on all of the displays are labeled according to callsigns. During the walkthrough, the TSD and Timeline displays were provided on individual slides to script the sequence of tasking as each scenario progressed. SMEs were given the option to return to prior steps and corresponding displays should they desire to revisit any specific information while proceeding through each use case.

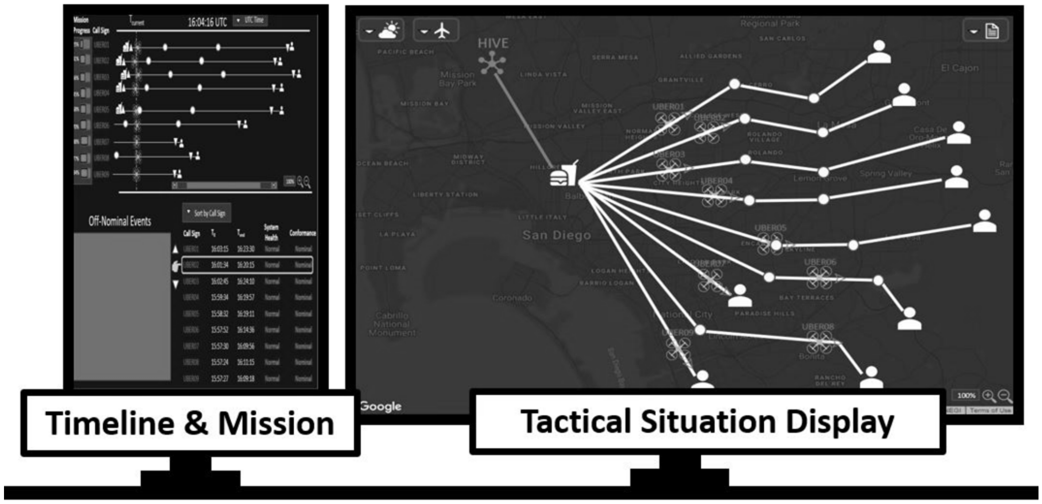


**Fig. 1 Ground Control Station.**

3. *Use Case Scenarios*

Four different use case scenarios were developed, each involving different combinations of operational and/or vehicular contingencies. In each scenario, a single operator managed nine sUAS, which all engaged in short-duration, food delivery missions. Although the work described below represents a vehicle configuration of 1:9, it is a crucial step in an effort of examining and scaling inputs and displays for M:N operations. Following an initial No Contingency scenario (i.e., a "nominal" operation), SMEs were tasked with confronting increasingly complex contingency events: The Simple Contingency scenario introduced a single-vehicle battery failure. The Complex Contingency scenario introduced a UVR, which is a planned or sudden restriction of airspace within UTM. The onset of the UVR affected multiple aircraft within several minutes of its issuance. Lastly, the Compound Contingency scenario combined the battery failure with the multiple vehicle UVR. Assumed mitigation strategies included an automatic transfer of power to an ancillary battery as well as the options to return-to-base (RTB), re-route, and modify waypoints. Table 1 provides a brief description of the four difference scenarios, including the number of affected UAS and number of task steps associated.

**Table 1. Contingency Scenario Summaries.**

| Use Case Scenario | Contingency Type | Contingency Mitigation | # of Affected UAS | # of Steps |
|---|---|---|---|---|
| No Contingency | N/A | N/A | N/A | 3 |
| Simple Contingency | Battery failure | Auto backup battery | 1 | 5 |
| Complex Contingency | UVR | Reroute, RTB, waypoint mod | 4 | 4 |
| Compound Contingency | UVR + battery failure | Auto backup battery, reroute, RTB, waypoint mod | 5 | 8 |

## C. Procedures

Informed consent forms were first emailed to SMEs who answered, signed, and returned them to the researchers. Because this process was performed remotely, these forms were designed so that the SMEs could complete them digitally. The cognitive walkthrough involved separately scheduled appointments with SMEs and began with an overview of the study as well as discussing expectations and logistics surrounding the remote nature of data collection. Demographics were also collected during this time. Voice and screen recording started after demographics were collected and following SME consent for recording. A presentation given by the research team introduced SMEs to project goals, proposed ConOps, the assumed area of operation, and the GCS configuration that would be explored. Once SMEs understood all of these elements, they were led through the use case scenarios.

Each use case began with a primer paragraph intended to prepare SMEs for the scenario conditions and the cognitive walkthrough process. Following the primer, the SMEs were guided through the use case scenarios in an unhurried and accommodating pace; each slide represented a display, either the TSD or Timeline display, and was presented based on where an operator was expected to look during the task flow. Each demonstrated an action that represented how an operator was expected to perform under those conditions. An example of the sequence for the No Contingency use case is pictured below in Fig. 2. The number of the slides depended on the complexity of the contingencies; because the use cases became more elaborate as the SMEs progressed, more slides were often required to document the steps required to resolve the associated contingencies. Each use case scenario ended following the resolution of the contingency event.

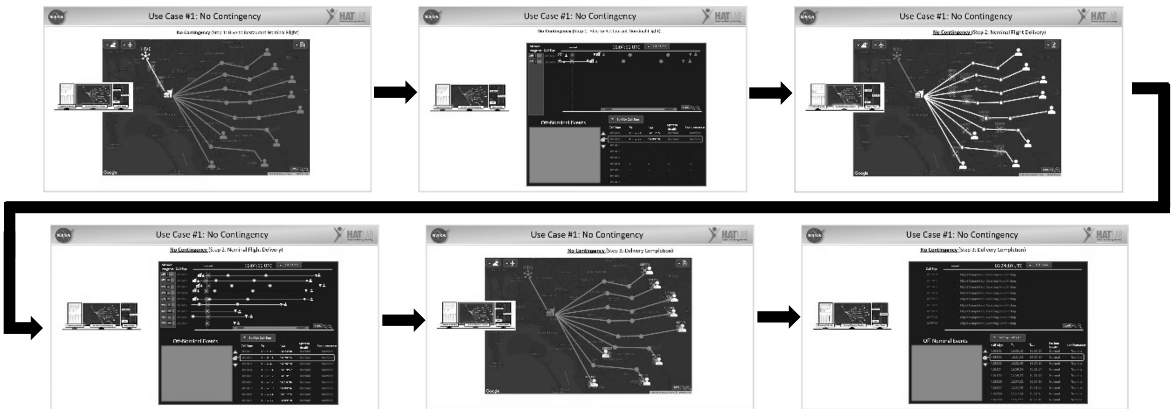


**Fig. 2 Use case slide sequence example: No Contingency.**

Following the briefing of a use case, SMEs were presented with questionnaires. In addition to evaluating the designs of the displays and controls, other topics that were explored with SMEs included potential types/functions of automation, the perceived time criticality of given scenarios/contingencies, and the perceptions of risk involved in contingency events along with the potential resolutions. A final portion allowed SMEs to answer additional questions as well as provide feedback on those questions — assessing how beneficial they would be if presented to SMEs of future 1:N and M:N studies. Lastly, SMEs were debriefed and the voice and screen recordings were stopped. The length of each interview depended on the amount of feedback provided from the SMEs; the average duration was approximately 90 minutes.

## III. Measures

### A. Ratings

1. *Risk*

The first measure, presented after every scenario, solicited SMEs' perception of risk; these were collected through ratings on five-point, likelihood and severity scales. These scales were modified from recommendations developed by Barr and colleagues [17]; an example is shown in Fig. 3. In order, risk ratings were delivered as the likelihood of the scenario occurring in the first place, the severity and likelihood if the scenario was successfully mitigated, and the severity and likelihood if the scenario was not successfully mitigated.

| | Category | Description |
|---|---|---|
| Severity if the Scenario is not Mitigated | (5) Catastrophic | Multiple Fatalities |
| | (4) Hazardous | Single Fatality and/or Multiple Serious Injuries |
| | (3) Major | Non-Serious Injuries |
| | (2) Minor | Inducing Fear |
| | (1) None | No safety effect for this scenario |
| Likelihood that the Scenario is not Mitigated | (5) Frequent | Will occur routinely |
| | (4) Probable | Will occur multiple times |
| | (3) Remote | Likely to occur once |
| | (2) Extremely Remote | Unlikely, but possible to occur |
| | (1) Extremely Improbable | Is expected to never occur |

**Fig. 3 Example of risk scale: Scenario is not mitigated.**

2. *Time Criticality & Automation Needed*

Following the risk scales, SMEs were asked their opinions of the time criticality and automation required for each scenario. Levels of time criticality were decided from discussions of the research team, and levels of automation were modified from the National Highway Traffic Safety Administration [18] and Parasuraman and colleagues [19]. The ratings themselves are provided below:

- **Time Criticality.** For this scenario, what level of action is required?
  - **Urgent:** Immediate action required
  - **Necessary:** Action needed soon
  - **Eventual:** Action needed eventually
  - **Potential:** Action potentially needed eventually
  - **None:** No action required
- **Automation Needed.** For this scenario, what level of automation should be used (i.e., that you are the most comfortable with)?
  - **Full Automation:** Vehicle is capable of performing all functions under all conditions. Operator has the option to control the vehicle.
  - **High Automation:** Vehicle is capable of performing all functions under certain conditions. Operator has the option to control the vehicle.
  - **Conditional Automation:** Operator is necessary, but not required to monitor the environment. Operator must be ready to take control at all times.
  - **Partial Automation:** Vehicle has combined automation functions, but operator must remain engaged with the task and monitor environment at all times.
  - **Operator Assistance:** Vehicle is controlled by the operator; but some assist features may be included in the vehicle design.
  - **No Automation:** Zero autonomy; the operator performs all tasks.

3. *Meaningful Human Control*

During a time of increased automation, the amount of control that is still afforded to human operators is often a source of contention [20]. The types of controls that currently are, as well as will be, enacted by humans will need to be safe, convenient, and effective. In an attempt to capture these considerations, a meaningful human control (MHC) questionnaire was developed that used seven-point scales, ranging from "Not at all" (1) to "Absolutely" (7), to measure the six dimensions detailed below; it also provided the open-ended question: "What does Meaningful Human Control mean to you?" SMEs in this study did not actually control the system, and therefore they were asked to answer the questions as if they had direct control of the vehicles under test.

- **Range of Options:** Did you have the range of response options required to respond as needed?
- **Temporal Availability:** Did you have the time to assess the situation and respond as required?
- **Interface Layout:** Did interface elements support an efficient and effective workflow?
- **Information Availability:** Was the information that you needed to respond available?
- **Workload (Anticipated):** Was your workload low enough for you to respond appropriately?
- **Overall MHC:** Did you feel you were able to exert meaningful human control?

### B. Open-Ended Questions and Dialogue

In addition to the scales, SMEs were also asked open-ended questions after scenarios as well as after the walkthrough. These questions gauged participant opinion on aspects of safety, the GCS configuration (e.g., effectiveness), workload (e.g., how many aircraft can be reasonably managed), information availability (e.g., how much is too much or too little), R&R (e.g., for humans and automation), the acceptability of the use cases on a whole, and how the proposed system compares to current practices and real-world applications. Additional input was recorded, including elaborations to ratings and other general feedback during the walkthrough.

## IV. Results

A comparison of medians was used to assess the responses to the survey questions. These reveal the minimum and maximum responses that include any outliers. Information collected from open-ended questions and conversational dialog provided further elaboration.

### A. Hazards Analysis: Likelihoods, Severities, and Time Demands

Hazard data are shown in Table 2 below. Results showed that the Complex Contingency (UVR) was perceived by the SMEs to be the scenario most likely to occur (*Mdn* = 4, *IQR* = 4-4); it was also the most likely to be mitigated (*Mdn* = 5, *IQR* = 4-5) as shown in Fig. 4 below. SMEs felt that, if not mitigated, severity was worse for the scenarios that involved battery failures, specifically the Simple (*Mdn* = 4, *IQR* = 3-4) and Compound (*Mdn* = 4, *IQR* = 4-4) Contingencies. This was due to the dangers present should a UAV lose power and plummet to the ground. Although slightly higher in the Compound Contingency (*Mdn* = 5, *IQR* = 5-5), SMEs also reported that time demand for all of the scenarios were between urgent and necessary; this means that a reaction would certainly, versus potentially, be needed immediately or soon.

**Table 2. Frequency of Ratings per Metric and Contingency Type.**

| | Rating | Likelihood to Occur | Severity if Mitigated | Likelihood to be Mitigated | Severity if Unmitigated | Likelihood to be Unmitigated | Time-Criticality |
|---|---|---|---|---|---|---|---|
| **Simple Contingency (Battery + UVR)** | **1** | 0 | 6 | 0 | 0 | 3 | 0 |
| | **2** | 1 | 4 | 0 | 1 | 3 | 0 |
| | **3** | 0 | 0 | 3 | 2 | 2 | 1 |
| | **4** | 7 | 0 | 0 | 6 | 1 | 5 |
| | **5** | 2 | 0 | 7 | 1 | 1 | 4 |
| **Complex Contingency (UVR)** | **1** | 0 | 7 | 0 | 0 | 1 | 0 |
| | **2** | 0 | 3 | 0 | 3 | 6 | 0 |
| | **3** | 0 | 0 | 0 | 2 | 2 | 2 |
| | **4** | 8 | 0 | 4 | 5 | 0 | 4 |
| | **5** | 2 | 0 | 6 | 0 | 1 | 4 |
| **Compound Contingency (Battery)** | **1** | 1 | 3 | 0 | 0 | 1 | 0 |
| | **2** | 0 | 7 | 1 | 0 | 5 | 0 |
| | **3** | 2 | 0 | 0 | 2 | 1 | 0 |
| | **4** | 7 | 0 | 4 | 7 | 3 | 2 |
| | **5** | 0 | 0 | 5 | 1 | 0 | 8 |

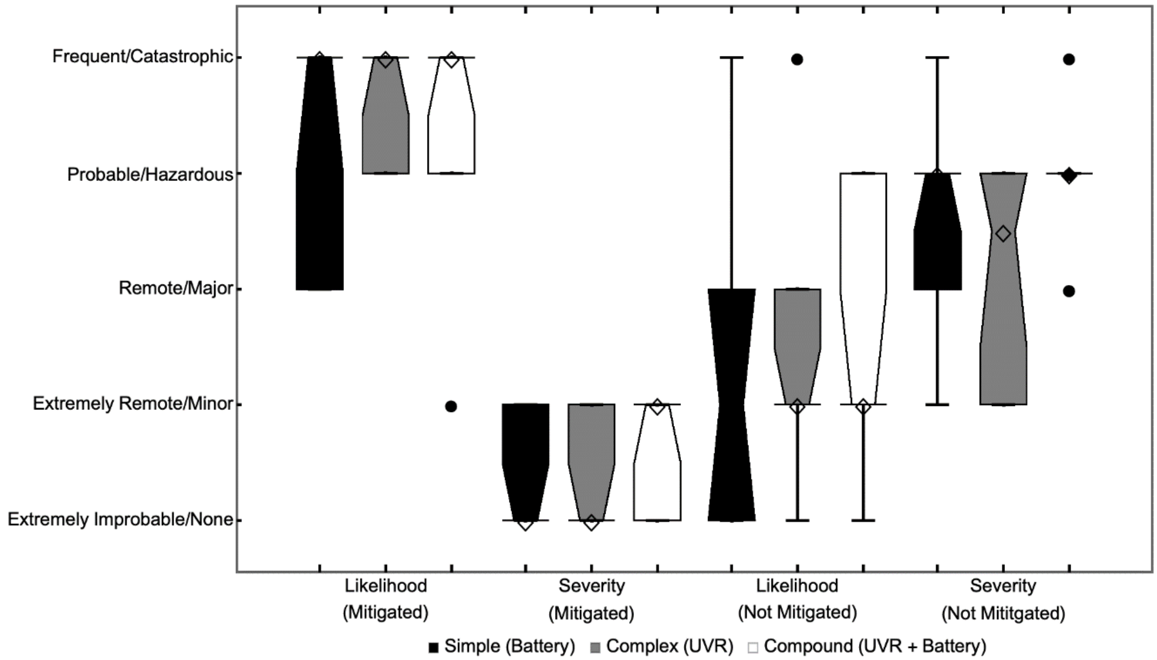


**Fig. 4 Distribution of risk rating scores, mitigated and not mitigated, by contingency. Filled circles = outliers, hollow diamonds = median values. Vertical axis labels = likelihood/severity scales, respectively.**

### B. Human Autonomy Teaming

On average, SMEs believed that automation should be conditional-to-high for all contingencies, meaning they preferred that the system governs most aspects of the flight while the operator is still actively involved and performing a supervisory role. Regarding meaningful human control (Fig. 5), SMEs rated the most agreeable aspects of MHC as Temporal Availability (*Mdn* = 6, *IQR* = 6-7) and Workload (*Mdn* = 6, *IQR* = 6-6). The lowest MHC scores were Information Availability (*Mdn* = 5, *IQR* = 4-6) and Overall MHC (*Mdn* = 5.5, *IQR* = 4-6); therefore, pilots, as a whole, believed that the proposed GCS configuration mostly satisfied the six dimensions of MHC asked by the questionnaire on the one ("Not at All") to seven ("Absolutely") scale.

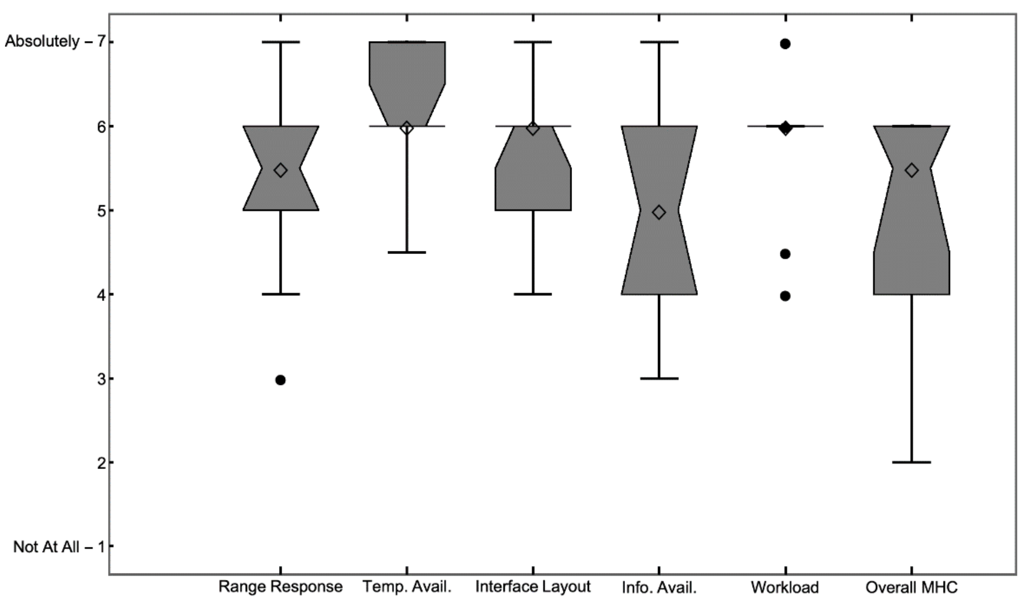


**Fig. 5 Distribution of Meaningful Human Control scores.**

### C. Emergent & Corroborated Themes

Because cognitive walkthroughs solicit information through a guided conversation, most feedback was derived from open and reciprocal dialog. These were not rated responses, but were instead provided as support for rated responses, answers to open-ended questions, or simply information shared by the SMEs as the thoughts entered their minds. The majority of identified themes related to hazards, GCS configurations, workloads, and duties and interactions of the automation and human operators. Most of the discussions regarding risk and hazards involved battery concerns; specifically, SMEs supported the concept that more information should be available about the battery as well as the onboard capability of an ancillary battery:

- "Information about the battery life… that probably would be my only other concern."
- "You're going to run into this battery problem pretty frequently I think."
- "So with an extra battery on board, that makes life a little easier."

GCS considerations often included display concerns like color schemes as well as the locations, clarity, accessibility, and quantity of the information presented. Reoccurring requests for additional data included weather, terrain, asset details (e.g., airspeed, altitude, health), and traffic (e.g., intruder) details. Pilots also felt that the GCS should afford more control to the operator in the forms of more flight options (e.g., changes in speed, altitude, and paths) as well contingency mitigation options (e.g., emergency landings as well as the options to stop/hover and perform all remedial actions simultaneously).

- "You can't have too much information."
- "I would want to be able to glance at the whole display and not have to spend my time trying to decipher each word."

Whereas SMEs believed that the workload for the use cases in the present study was reasonable, they also chose to discuss the problems that could arise if workload was too high or difficult to manage (e.g., stress and overload) as well as worries regarding low workload that may introduce boredom and complacency. These perceptions of workload depended on the amounts (i.e., singular versus compound) and types of contingencies (i.e., simple versus complex) that were occurring as well as the number of aircraft that were being managed. Pilots' opinions regarding the number of aircraft that could be successfully managed varied; these ranged from 5-20 assets with an average of 13.

Roles, responsibilities, and interactions between automation and the human operator was a popular topic captured in the majority of SME feedback. SMEs mostly believed that automation should be responsible for lower-level tasks like maintaining speed, altitude, stability, and power consistency. It was also agreed upon that the system should provide information and options, and execute those options when the humans are unable — such as in situations that require the enhanced computational or perceptual capabilities of the system or during operator incapacitation. For the human operator, SMEs largely supported the assistance of a secondary human operator: "I hope if things get bad I can pass off some of the aircraft to a fellow operator next to me." Additionally, SMEs believed that the meaningful duties of the operators should be managerial in nature; as the primary authority of all assets, they should monitor, approve, deny, override, and provide the final decisions for the actions and recommendations of automation:

- "I definitely see this is a managerial role."
- "If everything works perfectly well, all's you're doing is monitoring the system and ensuring that the UASs are going to the correct spots."
- "I would say to be in complete control of the aircraft with sufficient data."
- "It means the ability to monitor automated decisions, influence alternatives, and approve execution."
- "It's effectively a decision that is made in a dynamic environment by someone who's able to validate multiple, multiple variables and then adjust the automation to react appropriately.
- "As a pilot, if I'm going to make a decision, then potentially I'm going to be liable for the things going wrong. And if I'm unable to affect the outcome, then I have no meaningful control."

## V. Discussion and Conclusion

Collectively, SMEs found the scenarios to be realistic, and workload (e.g., number of aircraft as well as contingency management) to be reasonable given the display configurations and types of corrective actions that were proposed. SME feedback supported the assertion that the more commonly expected contingencies for operations like these are airspace restrictions and energy concerns for the UAVs. The latter, however, was discussed more frequently due to the dangers posed by aircraft falling to the ground (e.g., damage to people and property). SMEs believed that the GCS was a good starting point as an operator interface, but they also recommended that it should provide more information about assets and airspace, as well as provide more options for controlling aircraft, without also overloading an operator. The majority of responses discussed the duties and relationships between humans and automated systems. Specifically, automation controls basic, necessary flight functions, provides options to the human operator, and assumes command of controls when the human is unable to. Therefore, the human monitors that the system is behaving nominally and makes executive decisions based on the information from it.

SMEs provided precise and adamant suggestions regarding the topics discussed above, and this feedback revealed areas in need of further investigation. These topics were also explored in the present project, but feedback supported that they continue to be studied in future HITLs when displays become more mature: What specific types of information should be available? How much information is too much/too little? How and where should this information be displayed? How should visual, aural, and other distinguishing features be used for aircraft, routes, and alerts? Themes that were repeated among SMEs, despite not being intentionally pursued by the researchers, were backup human operators and inter-operator vehicle transfers: How could backup human operators assist during high-workload, contingency situations? How could procedures be optimized to allow for effective and efficient handoffs of multiple vehicles?

It should be noted that pilots were not directly in control of the scenarios and decisions that were being presented to them. This likely affected some of their rating scores as well as other general feedback. For example, pilots found that time availability and workload were agreeable within the use cases. This was because they were guided through the use cases at a leisurely pace while being shown remedial actions that would be performed by them in hypothetical situations. Therefore, they did not experience the time constraints or task difficulties associated with monitoring and controlling the system. In addition, lower Overall MHC scores may have been the result the simplified mockups as well as fact that SMEs were told how the system should be controlled instead of exercising that control themselves. This was supported by a follow-up survey study that demonstrated that, although the displays themselves are capable, effective, and allow for the progression of skill, the ways in which the SMEs encountered the displays did not support the satisfaction associated with choice [21]. Conversely, this perspective (i.e., observing the scenarios in a second-person, versus first-person way) could have given SMEs the opportunity to identify and recommend more information by not allowing portions of their attention to be absorbed in other cognitively-demanding, time-sensitive tasks.

By critically observing how the proposed tools affect specific contingency scenarios, SMEs provided a wealth of information about the strengths and weaknesses of the designs. Specifically, SMEs defined expectations of humans and system autonomy in delivery operations involving simultaneous control of multiple aircraft, the workloads and hazards of these operations, and the information and tools needed to manage the workloads and hazards. Although the designs were simplified to accommodate the remote nature of the study and elicit subjective feedback, they gathered information that allowed researchers to plan more mature displays for future experiments. For example, some results of this study, like the visual and informational aspects of the TSD and Timeline displays, were incorporated into designs for a later remote HITL study performed in Spring 2021. Additional studies for 2022 are being developed that continue to explore procedures for inter-operator vehicle transfers as well as other HAT tools and technologies.

## References


[1] Keane, J. F. and Carr, S. S., "A Brief History of Early Unmanned Aircraft," *Johns Hopkins APL Technical Digest*, Vol. 32, No. 3, 2013, pp. 558–571, URL: https://www.jhuapl.edu/Content/techdigest/pdf/V32-N03/32-03-Keane.pdf [retrieved 6 June 2021].

[2] Federal Aviation Administration, "Concept of Operations v2.0, Unmanned Aircraft System (UAS) Traffic Management (UTM)," FAA, 2020, URL: https://www.faa.gov/uas/research_development/traffic_management/media/UTM_ConOps_v2.pdf [retrieved 6 June 2021].

[3] Monk, K. J., Rorie, R. C., Sadler, G. G., Brandt, S., and Roberts, Z. S., "A Detect and Avoid System in the Context of Multiple-Unmanned Aircraft Systems Operations," *AIAA Aviation 2019 Forum*, Dallas, Texas, 2019. doi: 10.2514/6.2019-3315

[4] Federal Aviation Administration, "FAA Aerospace Forecast, Fiscal Years 2019-2039," FAA, 2019, URL: https://www.faa.gov/data_research/aviation/aerospace_forecasts/media/FY2019-39_FAA_Aerospace_Forecast.pdf [retrieved 6 June 2021].

[5] Fern, L., Kenny, C., Shively, R. J., and Johnson, W., "UAS Integration into the NAS: An Examination of Baseline Compliance in the Current Airspace System," *Human Factors and Ergonomics Society 56th Annual Meeting*, Vol. 56, No. 1, Boston, MA, 2012, pp. 41–45.
doi: 10.1177/1071181312561029

[6] Fern, L., and Shively, R. J., "A Comparison of Varying Levels of Automation on the Supervisory Control of Multiple UASs," *AUVSI's Unmanned Systems North America*, Washington, D.C., 2009, URL: https://www.researchgate.net/profile/Lisa-Fern-2/publication/257645042_A_COMPARISON_OF_VARYING_LEVELS_OF_AUTOMATION_ON_THE_SUPERVISORY_CONTROL_OF_MULTIPLE_UASs/links/004635258b5a8ed6b7000000/A-COMPARISON-OF-VARYING-LEVELS-OF-AUTOMATION-ON-THE-SUPERVISORY-CONTROL-OF-MULTIPLE-UASs.pdf [retrieved 6 June 2021].

[7] Cummings, M. L., Nehme, C. E., Crandall, J., and Mitchell, P., "Predicting Operator Capacity for Supervisory Control of Multiple UAVs," In *Innovations in Intelligent Machines-1*, Springer Berlin Heidelberg, 2007, pp. 11–37.

[8] Miller, C. A., and Parasuraman, R. "Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 49, No. 1, 2007, pp. 57–75.
doi:10.1518/001872007779598037

[9] Nehme, C. E., Scott, S. D., Cummings, M. L., and Furusho, C. Y., "Generating Requirements for Futuristic Hetrogenous Unmanned Systems," *Human Factors and Ergonomics Society Annual Meeting*, Vol. 50, No. 3, 2006, pp. 235–239.
doi:10.1177/154193120605000306

[10] Chen, J. Y. C., Barnes, M., and Harper-Sciarini, M, "Supervisory Control of Multiple Robots: Human-Performance Issues and User-Interface Design," *IEEE Transactions on Systems Man and Cybernetics Part C (Applications and Reviews)*, Vol. 41, No. 4, 2011, pp. 435–454.
doi:10.1109/TSMCC.2010.2056682

[11] Draper, M., Calhoun, G., Ruff, H., Mullins, B., Lefebvre, A., Ayala, A., and Wright, N., "Transition Display Aid for Changing Camera Views in UAV Operations," *1st Humans Operating Unmanned Systems (HUMOUS'08)*, Brest, France, 2008.

[12] Rubinstein, J. S., Meyer, D. E., and Evans, J. E., "Executive Control of Cognitive Processes in Task Switching," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 27, No. 4, 2001, pp. 763–797.
doi: 10.1037//0096-1523.27.4.763

[13] Monsell, S., Sumner, P., and Waters, H. "Task-Set Reconfiguration with Predictable and Unpredictable Task Switches," *Memory & Cognition*, Vol. 31, No. 3, 2003, pp. 327–342.
doi: 10.3758/bf03194391

[14] Porat, T., Oron-Gilad, T., Rottem-Hovev, M., and Silbiger, J., "Supervising and Controlling Unmanned Systems: A Multi-Phase Study with Subject Matter Experts," Frontiers in Psychology, Vol. 7, No. 568, 2016.
doi: 10.3389/fpsyg.2016.00568

[15] Mahatody, T., Sagar, M., and Kolski, C., "State of the Art on the Cognitive Walkthrough Method, Its Variants and Evolutions, International Journal of Human-Computer Interaction," *International Journal of Human–Computer Interaction*, Vol. 2, No. 8, 2010, pp. 741–785.
doi: 10.1080/10447311003781409

[16] Charmaz, K. and Belgrave, L., "Qualitative Interviewing and Grounded Theory Analysis," In *The SAGE handbook of interview research: The complexity of the craft*, 2012, pp. 347-366.
doi: 10.4135/9781452218403

[17] Barr, L., Newman, R., Ancel, E., Belcastro, C., Foster, J., Evans, J., and Klyde, D., "Preliminary Risk Assessment for Small Unmanned Aircraft Systems," *17th AIAA Aviation Technology, Integration, and Operations Conference*, Denver, CO, 2017.
doi: 10.2514/6.2017-3272

[18] National Highway Traffic Safety Administration, "Automated Vehicles for Safety," *NHTSA*, URL: https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety [retrieved 6 June 2021].

[19] Parasuraman, R., Sheridan, T., and Wickens, C., "A Model for Types and Levels of Human Interaction with Automation," *IEEE Transactions on Systems, Man, AND Cybernetics—Part A: Systems and Humans*, Vol. 30, No. 3, 2000, pp. 286–297.
doi: 10.1109/3468.844354

[20] Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G.K., Khlaaf, H., Yang, J., Toner, H., Fong, et al., "Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims," *Cornell University*, 2020.
arXiv:2004.07213v2

[21] Viramontes, J., "Display Considerations for Multi-UAS Operations," California State University, Long Beach, CA, 2021 (unpublished).