# Human-Autonomy Teaming Assistant to Support Small Uncrewed Aircraft Systems for Wildland Firefighting Operations

Victoria Dulchinos
*NASA Ames*
Moffett Field, CA, USA
victoria.l.dulchinos@nasa.gov

Jillian N. Keeler
*NASA Ames*
Moffett Field, CA, USA
jillian.n.keeler@nasa.gov

Garrett Sadler
*NASA Ames*
Moffett Field, CA, USA
garrett.g.sadler@nasa.gov

Linnea J. Holm
*SJSURF*
Moffett Field, CA, USA
linnea.j.holm@nasa.gov

Krish Pradhan
*SJSURF*
Moffett Field, CA, USA
krish.pradhan@nasa.gov

Vernol Battiste
*SJSURF*
Moffett Field, CA, USA
vernol.battiste@nasa.gov

Joel Lachter
*NASA Ames*
Moffett Field, CA, USA
joel.lachter@nasa.gov

Summer L. Brandt
*NASA Ames*
Moffett Field, CA, USA
summer.l.brandt@nasa.gov

***Abstract*-An exploratory human-in-the-loop simulation was conducted to investigate and characterize a Human-Autonomy Teaming (HAT) Assistant to support a remote operator of multiple small Uncrewed Aircraft Systems (sUAS) using a ground control station (GCS) in the context of a wildland fire surveillance mission. Operator performance using the GCS with the HAT Assistant (Assisted Mode) was compared to operator performance using the GCS without the HAT Assistant (Unassisted Mode) during two types of contingency-event scenarios (Low and High Complexity). In the Assisted Mode, the HAT Assistant provided updates to the level of risk to the mission along with recommendations for risk mitigation, which were not provided in the Unassisted Mode. No significant differences in objective performance and subjective ratings of workload, situation awareness, and trust in automation between the Assisted and Unassisted Modes were detected, however there were indications that participants preferred the Assisted GCS over the Unassisted GCS and directions for further development were explored. Additional work is necessary to further refine the HAT Assistant and better characterize its effects on remote operator performance while managing multiple sUAS assets. Future work is recommended to optimize the implementation of an assistant to support operator performance during different missions and across vehicle classes.**

***Keywords—Human-Autonomy Teaming (HAT), Automated Assistant, In-time Aviation Safety Management System (IASMS), Wildland Firefighting, Multi-vehicle Control, Ground Control Station (GCS), Mission Risk, Workload, Situation Awareness, Trust.***

## I. INTRODUCTION

This paper describes a human-in-the-loop (HITL) simulation exploring a preliminary implementation of a Human-Autonomy Teaming (HAT) Assistant designed to provide valuable information about changes to mission risk in the context of multi-vehicle control of small Uncrewed Aircraft Systems (sUAS) in a wildland firefighting operation. The primary goal of the exploratory study was to support development of this assistant by characterizing remote operator performance while managing multiple sUAS vehicles, with and without a HAT Assistant.

Wildland fires are increasing in size, severity and frequency with profoundly adverse impacts on society and the environment, see [1], [2] for an overview. Factors such as increasing temperatures due to climate change, excessive fuel loads, expanding fire season length, as well as overall changes to the wildland-urban fire interface exacerbate wildland fire risks [1] - [4]. NASA, in collaboration with the U.S. Forest Service, has conducted a series of workshops to understand the state-of-the-art, needs, and opportunities to improve wildfire management [5] - [7]. From these workshops, several safety challenges faced by the wildland firefighting community have been noted, which include, but are not limited to, a lack of persistent surveillance for fire detection and tracking; a lack of persistent aerial operations, particularly under poor visibility; a lack of technology to enable multiple types of aircraft operating simultaneously; and a lack of timely access to data for safety critical decision-making [7].

As highlighted in several of these workshops, aerial firefighting assets are effective at allowing fire crews to access remote locations, gather information, secure fire perimeters, save structures, and reduce fire intensity; however, in 2022, aviation accidents tied with medical events as the leading causes of wildland firefighting fatalities [8]. The use of sUAS is often considered a means to reduce operational risk but is not without cost. Integrating sUAS with existing aerial operations remains a challenge in part due to complex flight paths required for firefighting missions with minimal airspace management in a shared airspace [9], accordingly sUAS continue to be categorized as an airspace hazard by the National Wildfire Coordinating Group [10].

Challenges of integration notwithstanding, UAS, both small and large, are a valuable tool in fire detection, monitoring, and risk assessment and can extend flight operations to adverse conditions such as nighttime and heavy smoke. UAS technologies have seen important progress, including the development of smaller, more agile aircraft, with increased flight time, greater affordability, and better computational capabilities. As a result, their use has been extended to a wide range of applications. UAS have demonstrated utility for real-time surveillance and monitoring of active wildland fires as well as post-fire evaluation and recovery (e.g., acreage burned and reseeding maps [11], [12]). The question then becomes: how to utilize

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

the technological advantages and improvements offered by UAS without in turn introducing or exacerbating risk to firefighters and their mission?

Research within NASA's Aeronautics Research Mission Directorate aims to improve the effectiveness and safety of wildland fire management through the modernization of aerial operations in collaboration with the wildland firefighting community. To this end, areas of interest have emerged: (a) testing and evaluating safety-oriented capabilities to enable safe operational integration of UAS vehicles and (b) enhancing situational awareness and safety through common operating pictures. For UAS to provide a safer way to carry out operations—i.e., transferring risk from people to uncrewed systems—significant advances in technology and increases in automation will be required to control and manage aircraft. While current sUAS wildland fire operations requires a team of operators, a future concept of operation envisions semi-autonomous flight with the human on-the-loop or over-the-loop and allowing for multi-vehicle control by a single operator [9], [13], and [14].

Although automation can significantly reduce operator workload, it can also produce unexpected consequences. When automated systems encounter conditions that fall outside their design envelope, they sometimes experience brittle failures, characterized by a sudden transition from normal operation to failure. Human operators are often the fallback to correct such contingencies, but operators can find it difficult to understand the situation, because use of automation has taken them out of the loop (see [15]). Advances in automation may enable more capable systems that could assume functions typically assigned to human operators. Consequently, these systems would require decreased human intervention overall. However, this often has the side effect of taking people further out of the loop, making it harder for them to correct the problems that do occur [15]. Further, human supervision, direction, and cooperation may still be required, necessitating careful considerations for how to ensure safety and design assurance of such increasingly autonomous systems. Together, these issues pose a key barrier to the adoption of automation and thus to realizing the many benefits of UAS in this operational context.

To address these and other risks for new and evolving paradigmns, the National Academies recommended an In-Time Aviation Safety Management System (IASMS) that would focus on integrating real-time risk monitoring, assessment, and mitigation with a more responsive timeframe for detecting known risks and identifying emergent risks [16]. Given that humans are the most flexible problem solvers currently available, they are likely to remain a major part of any risk mitigation strategy for the foreseeable future. However, how humans interact with increasingly autonomous systems is a major challenge facing safety in design and operational assurance [14]. The current research seeks to support these evolving roles by incorporating human-automation teaming principles into the operation [17 - 18]. NASA research has shown the benefits of several principles of human-automation teaming, including: (a) improving transparency, so operators have better mental models of how the automation works [19]; (b) developing two-way communications between the human operator and automation so that both the automation and human can make use of information available to the other; and (c) maintaining meaningful human control where the human can override the increasingly autonomous system [20] - [22]. NASA formed a human-autonomy teaming committee which provided additional research recommendations centered around how to design responsible and collaborative automation and increasingly autonomous systems [18].

The ground control station (GCS) used in this study has been developed with these principles in mind (see [19] – [21], and [23] – [24] for a description). In particular, interaction with the GCS uses "plays" (as inspired by work from [25]). A play incapsulates the steps required to achieve a specific goal along with roles and responsibilities for carrying out those steps. This provides a language for the operator to inform the automation of goals and to see the goals the automation is working toward, improving communication, while allowing the operator to maintain meaningful control. Previous work with plays examined varying levels of automation according to the context and the level of trust in the automation [22]. This can create a path for safely automating a system by shifting the level of automation of certain steps in a play as trust in the automation increases. The HAT Assistant proposed in this paper strengthens the bi-directional communication in this system by creating a means for automation to direct operator attention to developing risks and to suggest plays that might mitigate them in a timely, safe, and efficient manner.

The research reported here is an initial investigation of the HAT Assistant in the context of multi-vehicle control of sUAS in a wildland firefighting operation. In particular, the ability of a remote pilot in command (RPIC) to manage four sUAS on a transect mission to map the progress of bulldozer lines and the fire progression were tested across two different GCS Modes (Assisted and Unassisted) with two levels of Complexity (Low and High). Hypothesis testing was not the primary objective of this effort, rather the focus was on the exploration of metrics and methods to be applied in follow-on research. This initial exploration characterizes and investigates the effect of the HAT Assistant on participants' measured performance, workload, situation awareness, and trust in the automation. Open-ended, qualitative feedback on tasks, procedures, and software interfaces will be used to inform the further development of the HAT Assistant.

## II. EXPERIMENTAL DESIGN

For this study we employed a within-subjects 2 × 2 (GCS Mode × Complexity) factorial design with two levels of GCS Mode ("Unassisted" and "Assisted") provided by the HAT Assistant, and two levels of Complexity ("Low" and "High") of the contingency event in the simulated scenario.

## III. PARTICIPANTS

Our participants were 10 Part 107 certificated pilots with an average age of 30.20 years (*SE* = 3.26 years), of which the majority (*N* = 8) were male. For civilian uncrewed flight time, participants reported an average of 115 hours (*SE* = 51 hours). All but one participant had crewed flying experience, where the average civilian flight time was 2,190 hours (*SE* = 1,379 hours) amongst the participants. Seven of the participants were Instrument Flight Rules (IFR) rated, along with numerous additional ratings and certifications. No military flight time was reported.

### A. *Ground Control Station*

The GCS software used for this study was a substantially modified version of the Vigilant Spirit Control Station (VSCS) [26], originally developed by the U.S. Air Force Research Laboratory (AFRL). The GCS featured two monitors with mouse and keyboard input (Fig. 1). The monitors included a main display known as the Tactical Situation Display (TSD), which provided a moving map of the operational area, the assets under control, and basic inputs. The secondary display featured a timeline of current asset mission progression and other related information, as well as features that varied depending on the level of assistance per scenario.

In both GCS Modes, participants had access to a play library similar in concept to those used in [20, 23]. The play library allowed participants to explore and execute various action sets (i.e., Return to Base, Reconfigure Asset, and Emergency Land) that could mitigate simulated contingency events. For the purposes of this study, the play library employed a limited set of plays as a protoype of the concept. The Return to Base (RTB) play would send selected vehicle(s) back to the original base of operation. The Reconfigure Asset play would amend the route of a selected vehicle to include additional routing originally assigned to another vehicle. This would only be executed if a vehicle was unable to complete its mission route. Finally, the Emergency Land play would send selected vehicle(s) to a predesignated landing site to land immediately. Each of the plays could be explored and/or executed for individual, multiple, or all assets (see Fig. 2).

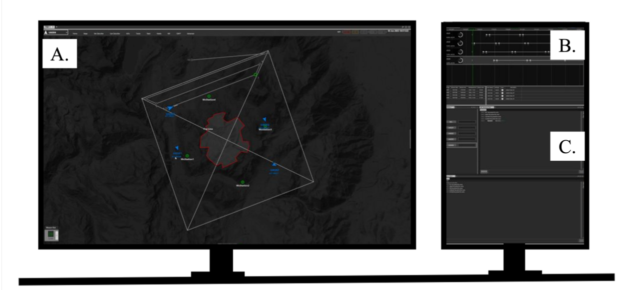


Fig 1. Ground control station setup, consisting of two displays, Tactical Situation Display (TSD, A.) on the left side and the Timeline (B.) HAT Assistant (C.) on the right side.

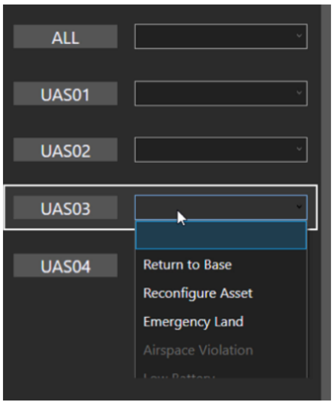


Fig 2. View of the play library available in both the Assisted and Unassisted versions of the VSCS.

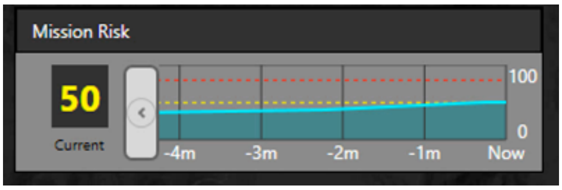


Fig 3. Expanded view of mission risk indicator available on the TSD.

In the Assisted Mode, participants had access to a HAT Assistant that communicated changes to mission risk and associated trends, as well as play recommendations to mitigate a contingency event. Mission risk was defined as the risk to completing the simulated mission successfully within the following parameters: (1) no loss of life or injury, (2) no loss of vehicle, (3) completed in the time allowed, and (4) with acceptable workload. The concept and associated values provided to the participants were an early effort to assign levels of risk to contingency events. A separate effort to further quantify and thoroughly define a mission risk scale is underway and out of scope of this paper.

Mission risk values were used to communicate estimated inherent risk and displayed on the TSD in the form of sparklines, representing historical changes to the risk value over the last four minutes (see Fig. 3). A color-coded numerical value representing current mission risk ranging from 0 to 100 was also provided. A green value represented low mission risk (i.e., 0-49), yellow moderate level of mission risk (i.e., 50-79), and red elevated mission risk (i.e., 80-100). In parallel, a chat message would be sent from the HAT Assistant indicating the new risk value and if the risk was increasing or decreasing from the last report. Participants could use the chat function to type "why" for additional information on factors influencing the change in risk or hover over the risk value on the mission risk indicator bar for the additional information. Changes to the mission risk score occurred concurrently as updates to the simulated contingency event(s) unfolded.

Additionally in the Assisted Mode, play recommendations were provided by the HAT Assistant when mission risk reached the moderate level threshold (i.e., a score of 50). The HAT Assistant communicated recommendations via the play library and Assistant chat window. Recommended plays were automatically preselected in the play library for the relevant assets (see Fig. 3). Upon receiving the recommendation, participants could view, modify, and/or execute the play, or explore other plays. In the chat window, participants could query why the play was recommended. Operational decisions were the responsibility of the participant.

In the Unassisted Mode, participants were responsible for recognizing the onset of simulated contingency event(s) and mitigating the situation by selecting plays from the play library. Plays could be previewed, modified, and executed in the same manner as in the Assisted Mode. Operational decisions were again the responsibility of the participant. Play recommendations, mission risk, and associated trends were not provided in the Unassisted Mode. All other aspects of the TSD and Timeline remained unchanged.

### B. *Training*

Prior to data collection, participants completed a consent form and background questionnaire. General GCS and mission training via power point and hands-on practice were provided first for all participants. Specific GCS Mode training (i.e., Assisted or Unassisted) and practice scenario were then provided with total morning training lasting about an hour before the first block of trials. After a lunch break, the next GCS Mode training and scenario were provided with total afternoon training lasting about a half hour before the second block of trials. The GCS Mode order was counterbalanced across participants.

### *C. Scenario*

The experimental scenarios were categorized by their scripted contingency event(s) as "Low Complexity" and "High Complexity." In the "Low Complexity" scenarios, a single asset experienced anomalously high current draw from the battery, significantly shortening estimated time to depletion. In the "High Complexity" event, all assets were affected by a weather inversion that unfolded over the simulation area. Participants each completed a total of four trials that were blocked by Assisted and Unassisted Modes (two blocks – one of each GCS Mode - each consisting of two trials – one of each Complexity). The order of the blocks was counterbalanced across the participants. Each trial lasted about 15 minutes.

## IV. METRICS AND ANALYSES

As an exploratory study, the metrics employed provide an initial characterization of how participants interacted with the HAT Assistant. This preliminary characterization will be used to guide future research and development of the HAT Assistant concept.

### *A. Subjective Metrics*

*Post-Trial*—After each trial, participants completed a post-trial questionnaire that compiled measurements of workload, situation awareness and trust in automation.

The NASA TLX measures subjective workload [27] and is composed of six dimensions, including (1) mental demand, (2) physical demand, (3) temporal demand, (4) performance, (5) effort, and (6) frustration. Participants rated items on 7-point Likert scales, ranging from 1 "Very Low" to 7 "Very High"; with the exception of the responses for performance ranging from 1 "Perfect" to 7 "Failure". A baseline for workload was captured prior to the experimental trials. Averages for each NASA TLX subscale and overall weighted scores were calculated following data collection.

The Situation Awareness Rating Technique (SART) measures situation awareness (SA) [29] across three domains (1) attentional supply, (2) attentional demand, and (3) understanding of the situation. Participants rated 10 items on a 7-point Likert scale ranging from 1 "Strongly Disagree" to 7 "Strongly Agree". A composite SART score was calculated with a maximum possible score of 46.

Merritt's Trust Scale measures trust in automation [29]. Participants rated six items on a 5-point Likert scale ranging from 1 "Strongly Disagree" to 5 "Strongly Agree" where higher numbers indicate more trust in the ground control station. A baseline measurement of trust in automation was administered prior to starting each block of trials.

*Post-Block*—After each block of trials (Assisted and Unassisted Modes), a post-block questionnaire was administered which included items regarding the usability (i.e., support for SA, efficiency, and effectiveness) and the capability to mitigate the effects of simulated contingency events with the two GCS Modes.

*Post-Simulation*—At the end of the day, a post-simulation questionnaire was administered which included general items about training, the GCS interface, and overall experience with the simulation. A semi-structured debrief was then conducted by the researchers. Participants provided verbal feedback on their decision-making process and general comments that may not have been captured in previous questionnaires.

### *B. Performance Metrics*

The following metrics were defined to capture participant response times to various events in the scenario (see Fig. 4):

- T1: Time between the onset of the contingency response (e.g., time when visibility dropped below Visual Flight Rules [VFR] minimums) and when the participant previewed a play from the play library.
- T2: Time between when the participant previewed a play from the play library to when the first play was executed. Note, in the High Complexity scenarios only one play was recommended.
- T3: Time between the first event (T1) and when the participant executed the first play (T2). This was the Total Response Time if only one play was executed.
- T4: If additional plays were previewed, time between when the participant previewed an additional play from the library and when the second play was executed. Note, in the Low Complexity scenarios two plays were recommended.
- T5: If additional plays were executed, time between the first event (T1) and when the second play was executed (T4). This was the Total Response Time if two plays were executed.

Additionally, participant behaviors such as the acceptance or rejection of recommendations by the HAT Assistant or whether participants executed a selected play of their own choosing were also documented.

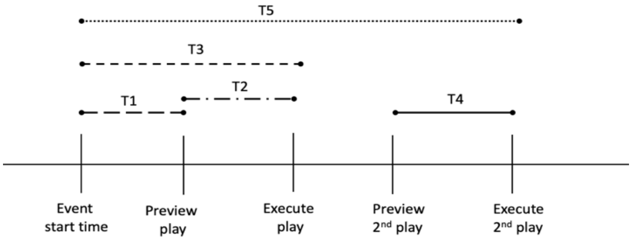


Fig 4. Event response timeline.

## V. RESULTS

### *A. Post-Trial Subjective Results*

After each trial, participants were given a standard battery of questions designed to measure their workload, situation awareness and trust in the system. None of these measures yielded significant differences across GCS Mode, Complexity, or the interaction of these factors.

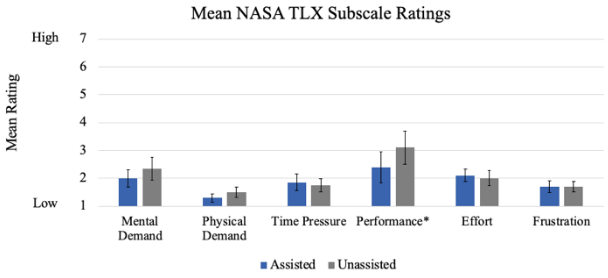
*Workload*—Overall workload ratings were low. Post-trial weighted NASA TLX ratings were similar for Assisted Mode ($M = 20.93$, $SE = 2.33$), Unassisted Mode ($M = 22.37$, $SE = 2.55$), Low Complexity ($M = 21.23$, $SE = 2.5$), and High Complexity ($M = 22.07$, $SE = 2.59$) scenarios.

See Fig. 5 for mean workload ratings by subscale. Note all ratings were low, though mean reported mental demand and effort averaged ratings of two or above compared to mean reported physical demand, temporal demand, or frustration level, where ratings averaged below two. Mean performance ratings were good.

*Situation Awareness*—Overall SA ratings were moderate. Composite SART scores were similar for Assisted Mode ($M = 27.25$, $SE = 2.47$), Unassisted Mode ($M = 27.40$, $SE = 2.02$), Low Complexity ($M = 27.40$, $SE = 2.10$), and High Complexity ($M = 27.25$, $SE = 2.06$) scenarios. See Fig. 6 for mean composite SART scores by condition. While mean SA ratings were consistent across all conditions, individual scores differed markedly from 15 to 41.

*Trust*—Overall participant trust in automation ratings were high. Aggregated ratings for Merritt's Trust Scale were similar for Assisted Mode (M = 4.55, SE = 0.22), Unassisted Mode (M = 4.33, SE = .25), Low Complexity (M = 4.43, SE = 0.24), and High Complexity (M = 4.44, SE = 0.21) scenarios.

See Fig. 7 for mean trust ratings by condition. On average, participants agreed to strongly agreed with trust in automation items with little variation in responses.



a. *Performance rating scale was 1 "Perfect" to 7 "Failure"

*Fig 5. Mean NASA TLX subscale ratings and SE bars in Assisted and Unassisted Modes*

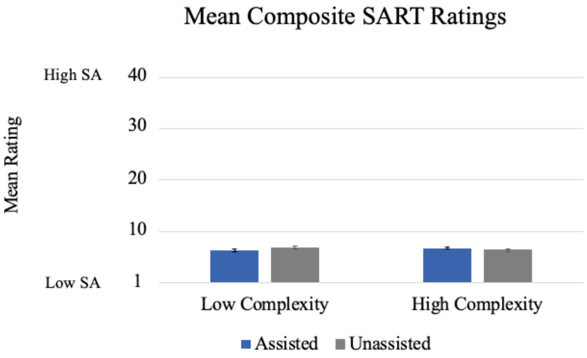


Fig 6. Mean SART ratings and SE bars in Assisted and Unassisted Modes for Low and High Complexity scenarios.

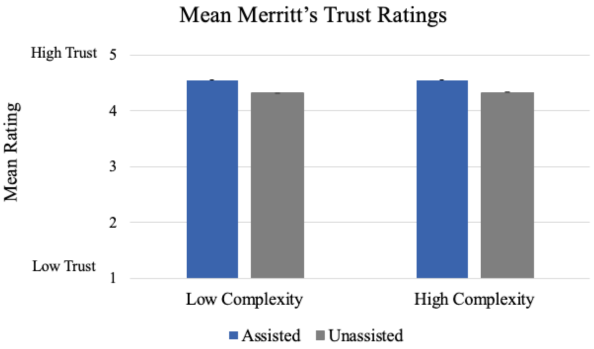


Fig 7. Mean Merritt's trust ratings and SE bars in Assisted and Unassisted Modes for Low and High Complexity scenarios.

### *B. Post-Block Subjective Results*

The post-block questionnaire contained seven items pertaining to usability, amount of information available, and capabilities provided by the GCS. The first four items asked participants to rank their agreement with statements regarding the efficiency and effectiveness of the GCS using a five-point Likert scale from 1 "Strongly Disagree" to 5 "Strongly Agree". On average, participants strongly agreed "*The GCS provided sufficient situation awareness to mitigate effects of unplanned events*" in the Assisted Mode ($M = 4.7$, $SE = 0.21$) and agreed in the Unassisted Mode ($M = 4$, $SE = 0.37$) and was the only analysis to yield a significant difference ($F(1, 9) = 5.4$, $p = 0.045$). See Fig. 8 for mean responses to the ranked statements. While participants agreed, on average, with the remaining statements, agreement was not significantly different between the GCS Modes. The GCS was found to provede sufficient ability to mitigate effects of unplanned events, to mitigate effectively, and with efficient interactions.

Another post-block item asked participants to select the sentence that reflected their opinion on the amount of information provided on the GCS displays. The three response options were:

- (1) "I needed more information to perform the tasks in the experimental trials,"
- (2) "The information provided in the display was necessary and sufficient to perform the tasks in the experimental trials," or
- (3) "There was more information than necessary to perform the tasks in the experimental trials."

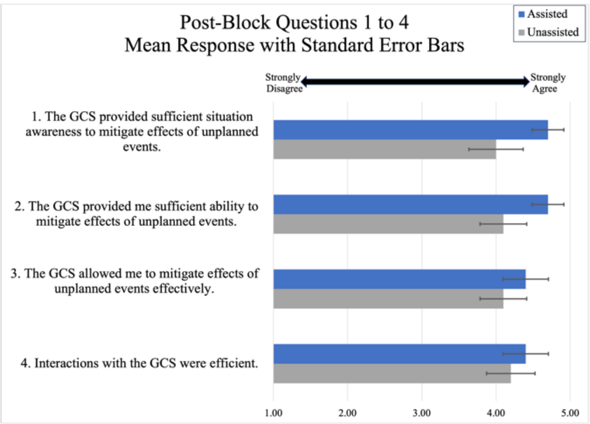


Fig 8. Post-Block Questions 1 to 4. Participant Mean Responses with standard error bars.

All 10 participants in the Assisted Mode selected option 2, indicating they felt the information was necessary and sufficient. However, there were mixed responses with the Unassisted Mode: 7/10 participants selected option 2, and 3/10 participants selections option 1, indicating they needed more information to perform the task.

Again participants were asked to select the sentence that reflected their opinion about the capabilities provided by the GCS. The three response options were:

- (1) "I needed additional capabilities to mitigate the unplanned events that occurred during the experimental trials,"
- (2) "The capabilities provided through the GCS were necessary and sufficient to mitigate the unplanned events that occurred during the experimental trials," or
- (3) "There were more capabilities than necessary to perform to mitigate the unplanned events that occurred during the experimental trials."

All 10 participants in the Assisted Mode selected option 2, indicating they felt the GCS capabilities were necessary and sufficient for handling the contingency events. However, in the Unassisted Mode, the results were mixed: 4/10 participants selected option 2, 4/10 participants selected option 1, and 2/10 participants selected option 3, indicating there were more capabilities than necessary.

The final post-block item, gave participants the option to provide open-ended comment on the GCS displays and interfaces. When examining the participants responses to this prompt, most comments referenced issues with the user interface and did not directly refer to the interactions with the HAT Assistant. In response to the Assisted Mode, a participant remarked: "*When hovering over the risk score, the message bar disappeared too quickly. I wish I had more control over that aspect to read and process the information*." For the Unassisted Mode, another participant stated: "*Although more monitoring was required, the GCS provided sufficient information to complete the task*."

### C. *Post-Simulation Results*

The post-simulation questionnaire asked a range of general questions about the overall experiment. Response options were on a 5-point Likert scale ranging from 1 "Strongly Disagree" to 5 "Strongly Agree." When asked about training, 9/10 participants strongly agreed they received sufficient training and 1/10 somewhat disagreed. When asked about the realism of the scenarios, 5/10 participants strongly agreed the scenarios were realistic, 4/10 somewhat agreed, and 1/10 neither agreed nor disagreed with the statement. When asked about the realism of the simulation, 6/10 participants rated the simulation as very realistic, 3/10 somewhat realistic, and 1/10 neither realistic nor unrealistic.

Additionally, participants were asked to rate whether the Unassisted and Assisted GCSs had all the features necessary to accomplish the mission. All 10 participants strongly agreed that the Assisted GCS provided all necessary features; however, three participants gave the Unassisted GCS lower ratings (2/10 somewhat agreed, 1/10 somewhat disagreed). Asked to elaborate on their answers, all three pointed to the utility of having the Assistant share the monitoring of aircraft. For example, one stated: "*Reminder from HAT made sure I didn't miss anything*." Similar sentiments were voiced by three additional participants who nonetheless strongly agreed that the Unassisted GCS had all the necessary features (e.g., "*More workload was put on the pilot since monitoring the weather and battery of the sUAS was to be done manually. Although, the GCS provided sufficient data to support the tasks*."). When asked if the HAT Assistant was useful for the wildland fire surveillance mission, 9/10 particpants strongly agreed and 1/10 somewhat agreed.

After all simulation trials were completed, a semi-structured debrief was conducted. Responses were varied. On the topic of useful information provided, some participants cited the weather information as it supported situation awareness. Still, others found the weather information to be the least useful due to the user interaction to toggle open or closed an information box on the TSD. The battery information on the timeline was called out by some as the most useful information. On the topic of required minimum capabilities, responses included the plays in the play library, the mission information provided in the mission chat, as well as the battery and weather information. Nine participants named the HAT Assistant as useful, eight of whom liked the HAT Assistant chat capability isolated from the chat capability with other mission personnel.

### D. *Performance Data Results*

*Behavioral Interactions*—For the Unassisted Mode with a Low Complexity event, 2/10 participants previewed and executed a play, with the majority failing to notice the power loss event unfolding. Additionally, only one of the two participants called a secondary play (i.e., Recongfigure Asset) to maximally resolve the battery loss event, with the other participant choosing only the Return to Base play. In the Unassisted Mode with a High Complexity event, a greater number of participants noticed the weather inversion and took action, with 5/10 chosing to preview and execute the Return to Base play for all assets.

For the Assisted Mode, both with Low and High Complexity events, 5/10 participants previewed a play before executing. Regardless of event complexity, 10/10 participants executed at least one play. All but one participant chose to follow the recommended play provided by the HAT Assistant in both Assisted scenarios; one participant took a more proactive approach and executed a desired play before the HAT Assistant provided a recommendation. For the Low Complexity event that included a secondary play, 9/10 participants previewed a secondary play (i.e., Reconfigure Asset) before executing. All participants executed a secondary play for the Low Complexity event in the Assisted version.

*Response Times*—Descriptive statistics were calculated for each response time measurement and are reported in Table 1 along with the respective number of participants who completed that action. Response times were generally longer in the Low Complexity scenarios, given that the desired actions were to call two separate plays to solve the situation (i.e., Return to Base and Reconfigure Asset plays), whereas the High Complexity scenario could be solved by using the "All" function to have all assets execute the Return to Base play. Most importantly, participants were generally faster and more consistent at responding to the presented contingency events in the Assisted Mode than the Unassisted Mode.

TABLE I. MEAN MEASURED RESPONSE TIMES BY CONDITION

| | **Overall** | | **Low Complexity** | | **High Complexity** | |
|---|---|---|---|---|---|---|
| **Response Time** | *Assisted* | *Unassisted* | *Assisted* | *Unassisted* | *Assisted* | *Unassisted* |
| **T1** | 18.17s<br>(1.30s)<br>$N = 2$ | 76.25s<br>(42.25s)<br>$N = 2$ | 24.75s<br>(3.52s)<br>$N = 3$ | 74.00s<br>(19s)<br>$N = 2$ | 17.00s<br>(2.99s)<br>$N = 4$ | 60.00s<br>(27.28s)<br>$N = 4$ |
| **T2** | 7.38s<br>(1.43s)<br>N = 3 | 14.50s<br>(9s)<br>$N = 2$ | 7.40s<br>(1.91s)<br>$N = 4$ | 9.50s<br>(2.50s)<br>$N = 2$ | 7.60s<br>(2.50s)<br>$N = 4$ | 12.80s<br>(5.83s)<br>$N = 4$ |
| **T3** | 38.89s<br>(8.30s)<br>N = 5 | 80.25s<br>(40.75s)<br>$N = 2$ | 43.00s<br>(15.53s)<br>$N = 5$ | 77.50s<br>(15.50s)<br>$N = 2$ | 33.50s<br>(11.57s)<br>$N = 10$ | 66.80s<br>(28.09s)<br>$N = 5$ |
| **T5** | 51.11s<br>(7.74s)<br>$N = 5$ | 84.25s<br>(36.75s)<br>$N = 2$ | 67.44s<br>(13.17s)<br>$N = 5$ | 85.50s<br>(7.50s)<br>$N = 2$ | - | - |

## VI. DISCUSSION

This study was an initial exploration of a HAT Assistant in the context of multi-vehicle control of sUAS in a wildland firefighting operation. Our goal was to characterize the utility of such an assistant to a remote operator managing multiple sUAS vehicles during a wildland fire surveillance mission. The overall outcome of this effort offers promise for future development of the assistant to support the remote operator managing multiple sUAS vehicles.

The results of the analysis of the post-trial subjective ratings of workload, situation awareness, and trust yielded no significant differences in ratings between the Assisted and Unassisted Modes. In regard to the workload results, it should be noted that there was also no significant difference between the High and Low Complexity conditions, suggesting that the High Complexity condition was not difficult enough to surface factors that might have reduced workload. With the situation awareness results, the presence of an assistant might have two opposing effects: the automation of risk tracking might reduce the attention operators need to place on tracking, say, battery levels or visibility, thus reducing situation awareness. At the same time, when the assistant alerts operators to the existance of risks, it likely increases situation awareness. Overall, the moderate ratings found in these results may be an indication that the assistant requires further development to reach a significant degree of value in providing situation awareness to the participant. It also should be considered that the SART measure may not have the level of sensitivity to detect differences and perhaps other methodologies should be explored to study situation awareness in this context. Ratings of trust were similar across all conditions. Even though instructions called out the HAT Assistant, it is unclear if participants were able to discern between ratings of trust of the HAT Assistant and ratings of trust of the overall ground control station. As with situation awareness, it is not clear whether one should expect the addition of automation to increase or decrease trust in a system. The current results suggest that the introduction of a basic assistant does not strongly affect either metric.

The post-block and post-simulation questionnaire results provided an additional layer of information beyond the results of the post-trial subjective data. The post-block questions assessed opinions and impressions regarding the general usability aspects of the displays and GCS. Some of these results point to a preference for using the GCS with the Assistant. For example, participants using the GCS in the Assisted Mode indicated stronger agreement with the statement that the GCS provided sufficient situation awareness to mitigate the effects of unplanned events than in the Unassisted Mode. Also, participants in the Assisted Mode unanimously agreed that the amount of information provided on the GCS displays was necessary and sufficient to perform the tasks in the experimental trials, while in the Unassisted Mode, the results were mixed with some participants indicating they needed more information to perform the tasks in the experiment trials. Additionally, when asked to provide an opinion about the capabilities of the GCS, participants in the Assisted Mode unanimously responded that the capabilities provided through the GCS were necessary and sufficient to mitigate the unplanned events that occurred during the experimental trials, while in the Unassisted Mode, the response was mixed with some participants indicating they needed additional capabilities, or that there were more capabilities than necessary. The post-simulation results also trended toward participant preference for the Assistant. All participants in the Assisted Mode strongly agreed that the Assisted GCS (i.e., with HAT Assistant) had all the features necessary to accomplish the mission. Furthermore, 9 out of 10 participants strongly agreed that "*The HAT Assistant was useful for wildland fire surveillance missions.*" Another aspect of the results that points to the benefit of the Assisted Mode can be seen by examining the performance data, where it was found that participants were generally faster and more consistent at responding to the presented contingency events in the Assisted Mode than the Unassisted Mode.

This study provided validation for some of the design decisions made in developing the HAT Assistant. Participants were generally positive about the decision to implement the assistant as a separate, chat-based agent. Some results indicate a general preference for the Assisted Mode. Other comments suggest modifications. For instance, several participants thought that the reasoning behind the assistant's recommendations should be presented without further prompting by the operator. Future studies of an Assistant with increased capabilities will focus on aiding sUAS operators during more complex high-tempo disaster response events, where multiple UAS vehicles could aid in providing logistics, search and rescue, and surveillance and reconnaissance missions. It is possible that the utility of the HAT Assistant will stand out in measures such as workload in these more complex environments. Data from these future studies will provide additional insight into improving the partnership between our operators and our HAT Assistant. The outcome of this exploratory study provides evidence to support future research efforts with the aim of conducting more rigorous hypothesis testing with a greater number of participants and a greater number of trials. Future research will assess the utility of the Assistant in aiding the remote operator managing multiple sUAS in the context of disaster response operations. Additionally, future work should endeavor to understand how to optimize human-autonomy teaming to create a resilient system under conditions where automation alone may perform poorly, and where the human operator may benefit from having the support of an assistant.

## ACKNOWLEDGMENT

This work was supported by the System-Wide Safety (SWS) project, under the Airspace Operations Program within the NASA Aeronautics Research Mission Directorate (ARMD). We would like to thank our pilot participants and human factors collegues who gave their time and feedback to help advance these concepts.

## REFERENCES

[1] Mueller, S. E., et al., "Climate relationships with increasing wildfire in the southwestern US from 1984 to 2015," in *Forest Ecology and Management*, 2020, vol. 460. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0378112719314744

[2] Richardson, D., et al., "Global increase in wildfire potential from compound fire weather and drought," in *npj Climate and Atmospheric Science*, 2022, vol. 5, no. 23. [Online]. Available: https://www.nature.com/articles/s41612-022-00248-4

[3] Radeloff, V. C., et al.,"Rapid growth of the US wildland-urban interface raises wildfire risk," in *Proceedings of the National Academy of Sciences*, 2018, vol. 115, no. 13, pp. 3314-3319. [Online]. Available: https://www.pnas.org/doi/10.1073/pnas.1718850115

[4] Tedim, F., et al., "Defining extreme wildfire events: difficulties, challenges, and impacts," in *Fire*, 2018, vol. 1, no. 9. [Online]. Available: https://www.mdpi.com/2571-6255/1/1/9

[5] Kopardekar, P., and Grindle, L. (2021). NASA ARMD wildfire management worksop. https://nari.arc.nasa.gov/tfrsac-wildfire

[6] NASA, "NASA Science Mission Directorate (SMD) wildfire stakeholder engagement workshop: summary and key findings," Draft Report, 2022. [Online] Available: https://nari.arc.nasa.gov/sites/default/files/attachments/NASA%20SMD%20Wildfire%20Stakeholder%20Workshop%20Summary_DRAFT.pdf

[7] Lehman, S. M., et al., "NASA system-wide safety wildland firefighting operations workshop," 2022, Rep. no. NASA/TM-2022001471. [Online] Available: https://ntrs.nasa.gov/citations/20220014721

[8] Wildfire Lessons Learned Center, "2022 incident review summary." 2022. [Online]. Available: https://wildfirelessons.blog/wp-content/uploads/2023/05/2022_LLC_Incident_Review_Summary.pdf

[9] Flight Safety Foundation, "Looking to the future: safety system needs for humanitarian UAS operations," 2021. [Online]. Available: https://flightsafety.org/wp-content/uploads/2021/07/D6-DMFR-FINAL.pdf

[10] NWCG, "NWCG standards for airspace coordination," Rep. no. PMS 520, July 2023. [Online]. Available: https://www.nwcg.gov/sites/default/files/publications/pms520.pdf

[11] Hinkley. E. A., & Zajkowski, T., "USDA Forest Service – NASA: unmanned aerial systems demonstrations – pushing the leading edge in fire mapping," in *Geocarto Internation,* 2011, vol. 26, no. 2, pp. 103-111. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/10106049.2011.555823

[12] Andrew Avitt, "Mechanized birds: Wildland firefighting with drone intelligence," US Forest Service, October 2022. [Online]. Available: https://www.fs.usda.gov/features/mechanized-birds-wildland-firefighting-drone-intelligence

[13] Ellis, K. K., et al., "A concept of operations (ConOps) of an in-time aviation safety management system (IASMS) for advanced air mobility (AAM)," in *AIAA Scitech 2021 Forum*, 2021, pp. 1978. [Online] Available: https://arc.aiaa.org/doi/abs/10.2514/6.2021-1978

[14] Prinzel III, L. J., et al., "Human interfaces and management of information (HIMI) challenges for "in-time" aviation safety management systems (IASMS)," in *HCI 2022: Human Interface and the Management of Information: Applications in Complex Technological Environments,*, 2022, pp. 367-387. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-06509-5_26

[15] Endsley, M. R., "From here to autonomy: lessons learned from human–automation research," in *Human Factors,* 2017, vol. 59, pp. 5–27. [Online]. Available: https://journals.sagepub.com/doi/10.1177/0018720816681350

[16] National Academies of Sciences, Engineering, and Medicine, "In-time aviation safety management: challenges and research for an evolving aviation system," Washington, DC, 2018. [Online]. Available: https://doi.org/10.17226/24962.

[17] Shively, R. J., Lachter, J., Koteskey, R., & Brandt, S. L., "Crew resource management for automated teammates (CRM-A)," in *HCII 2018,* pp. 215-229. [Online]. Available: https://dl.acm.org/doi/abs/10.1007/978-3-319-91122-9_19

[18] Holbrook, J. B., et al., "Enabling urban air mobility: Human-autonomy teaming resesearch challenges and recommendations," in *AIAA AVIATION Forum*, 2020, pp. 3250. [Online]. Available: https://humansystems.arc.nasa.gov/publications/6_2020_3250.pdf

[19] G. Sadler et al., "Effects of transparency on pilot trust and agreement in the autonomous constrained flight planner," in *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, 2016, pp. 1-9. [Online]. Available: https://ieeexplore.ieee.org/document/7777998

[20] Brandt, S., Lachter, J. B., Russel, R., and Shively, R. J., "A human-autonomy teaming approach for a flight-following task," in *AHFE 2017: Advances in Neuroergonomics and Cognitive Engineering*, 2017, vol. 586, pp. 12–22. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-60642-2_2

[21] Shively, R. J., et al., "Why human-autonomy teaming?," in *AHFE 2017: Advances in Neuroergonomics and Cognitive Engineering*, 2017, vol. 586, pp. 3-11. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-60642-2_1

[22] Mecacci, G., & Santoni de Sio, F., "Meaningful human control as reason-responsiveness: The case of dual-mode vehicles," in *Ethics and Information Technology*, 2020, vol. 22, pp. 103-115. [Online]. Available: https://link.springer.com/article/10.1007/s10676-019-09519-w

[23] Ho, N., et al., "Application of human-autonomy teaming to an advanced ground station for reduced crew operations," in *Proceedings of the 36th Digital Avionics Systems Conference*, 2017, vol. 3, pp. 99-128. [Online]. Available: https://link.springer.com/article/10.1007/s42454-021-00026-2

[24] Rorie, C., Battiste, V., & Shively, R. J., "m:N and human autonomy teaming concepts for high denisty vertiport operations," 2021, Rep. no. NASA/TM-20210025392. [Online]. Available: https://ntrs.nasa.gov/api/citations/20210025392/downloads/NASA_TM_20210025392.pdf

[25] Miller, C., et al., "The playbook TM approach to adaptive automation," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 49, pp. 15-19. [Online]. Available: https://doi.org/10.1177/154193120504900105

[26] Feitshans, G. L., Rowe, A. J., Davis, J. E., Holland, M., and Berger L., "Vigilant Spirit Control Station (VSCS): the face of COUNTER," August 2008, Rep. no. AFRL-RH-WP-TR-2012-0015. [Online]. Available: https://apps.dtic.mil/sti/tr/pdf/ADA594380.pdf

[27] Hart, S. G., “NASA-Task Load Index (NASA-TLX), 20 years later,” in *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, 2006, vol. 50, no. 9, pp. 904-908 [Online]. Available: https://journals.sagepub.com/doi/10.1177/154193120605000909

[28] Taylor, R. M., “Situational awareness rating technique (SART); The development of a tool for aircrew systems design,” in *Situational Awareness*, 1st ed. Routlege, 2017, pp. 111-128. [Online]. Available: https://www.taylorfrancis.com/chapters/edit/10.4324/9781315087924-8/situational-awareness-rating-technique-sart-development-tool-aircrew-systems-design-taylor

[29] Merritt, S. M., “Affective processing in human-automation interactions,” in *Human Factors,* 2011, vol. 53, no. 4*,* pp. 356-370. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/21901933/