Creating a Well-Situated Human-Autonomy Team:

The Effects of Team Structure

Elizabeth Marie Frost
Wright State Research Institute[1]
UNITED STATES OF AMERICA

Elizabeth.Frost.6.ctr@us.af.mil

Jessica Bartik **Gloria Calhoun** **Sarah Spriggs**
Air Force Research Laboratory
UNITED STATES OF AMERICA

Jessica.Bartik.1@us.af.mil Gloria.Calhoun@us.af.mil Sarah.Spriggs@us.af.mil

Heath Ruff **Kyle Behymer**
Infoscitex[1]
UNITED STATES OF AMERICA

Heath.Ruff.ctr@us.af.mil Kyle.Behymer.1.ctr@us.af.mil

ABSTRACT

Intelligent agent technologies are increasing the potential capacity for systems to behave more autonomously and are enabling more advanced human-autonomy teaming. The objective of this research was to gain a better understanding of future applications of human-autonomy teaming for the command and control of semi-autonomous systems. Specifically, the task division and relationships/roles between a human operator and an autonomous teammate were empirically examined. A control station that supports single operator management of multiple simulated unmanned vehicles performing a base defence mission was employed to study the implications of team structure and mission complexity on team performance. Operator-driven and role-driven team structures were evaluated across two levels of mission complexity by both human-human teams and human-autonomy teams.

1.0 INTRODUCTION

Effective teamwork is essential for achieving reliable operations in environments that are hypercomplex, dynamic, time compressed, and rely upon synchronized outcomes, such as military command and control [1]-[2]. Thus, the application of advanced intelligent agent capabilities to more complex and dynamic task environments, has led to an interest in research looking at how to support teaming between human operators and autonomous systems (e.g., [3]-[5]). With human-autonomy teaming being a relatively new concept, the majority of published teaming literature is focused on human-human teams. However, understanding what makes these

[1] This research was sponsored by the Air Force Research Laboratory (AFRL) under Contract FA8650-14-D-6500. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of AFRL or the US Government.

teams effective may provide insight for human-autonomy teams [6]. In the large body of research looking at different aspects of effective human teams in a wide range of domains, three specific factors have been shown to have significant influence on team performance: team cognition, communication/coordination, and team structure [7]-[9].

Team cognition, an emergent collection of cognitive processes, has been shown to be a driver of effective teams [10]-[11]. Team cognition spans various constructs such as shared mental models, transactive memory, and shared cognition, generally addressing both task knowledge and team knowledge. Task knowledge focuses on features of the team's tasks such as understanding the tools available, the task objective, and constraints. Team knowledge addresses the interaction between team members as well as the members of the team themselves, including their knowledge, skills, status, and intentions. Teammates with shared task and team mental models have an effect on overall team performance by positively influencing team processes [12]. For example, team members with a shared mental model are more likely to work towards a common objective and that eases coordination of their actions, as well as collaboration and communication [12].

Team cognition research has largely focused on human-human teams but it has been argued that the concept of shared mental models is also relevant for human-autonomy teams [13]-[14]. In this case it is important for the human operator to have appropriate system and teammate understanding in order to coordinate effectively with the autonomy [15]. This understanding can be supported by providing a human operator with sufficient training as well as experience working with the system in order to develop accurate task and team mental models. The system's interfaces will also need to support information sharing important to the development of team cognition. Additionally, the system will need to enable communication and coordination among the teammates, which is necessary for team cognition to have a positive influence on performance [12], [16].

The importance of effective communication and coordination to achieve successful teamwork is well-researched (e.g., [8], [17]-[18]). These two aspects naturally go hand in hand in that the ability to effectively communicate is vital for team coordination. This goes beyond just human-human teams, and will also need to be addressed in human-autonomy teams if they are to succeed [3], [5]. However, human-autonomy teams face different challenges when it comes to communication and coordination, such as the current lack of language technologies that support the types of communication that are regularly used in human-human teams when performing task management [19]. This means that communication that occurs naturally in human-human teams, may now present a design challenge for human-autonomy teams, making it difficult to support team performance without causing information overload and degrading information retrieval [20]-[21].

Though an interface can be designed to help support human-autonomy coordination and communication, the team structure for sharing task completion also needs to be considered. Team structure is commonly referred to as "the division of the team task into component pieces of information and capabilities and the assignment of these elements to individuals in the team" [22, p. 302]. Various human-human teams have been compared in terms of division of information and capabilities, with the results indicating that some team structures yield more superior performance than others [9]. In particular, human-human teaming research has shown that a non-hierarchical team structure may result in better team performance than a hierarchical team structure [23]. Current supervisory control approaches usually involve task delegation using a hierarchical structure, similar to a supervisor delegating tasks to a subordinate. Alternatively, a functional team structure, that involves teammates having defined roles while still working together collaboratively to achieve certain goals, is representative of a non-hierarchical structure. However, it is important to note that the task environment can influence the effectiveness of various team structures [24]. For example, in a review of air traffic control systems it was observed that having a hierarchical structure during normal operation times (with low to moderate demand) allowed for a controlled environment that still provided adequate responses. In high stressed situations, though, a shift towards a flatter, decentralized structure allowed for the adaptability and flexibility needed to respond successfully.

2.0 OBJECTIVE

The present research specifically examined two team structures, operator-driven (supervisory control) structure compared to a role-driven (functional) one. These two team structures were examined in both human-human teams and human-autonomy teams, with each type of team completing the same mission tasks. In the operator-driven team structure, the human operator received all the tasking and then decided which tasks to delegate to an autonomous or human teammate and which tasks to personally complete. Alternatively, in the role-driven structure, each team member fulfilled roles based on their expertise and skills (with the human teammate's skills mimicking that of the autonomous teammate). The role-driven team structure was less centralized, such that the teammate automatically began completing tasks without waiting for tasks to be delegated. Experimental participants worked with their respective teammate (either human or autonomous) to complete base defence mission tasks using multiple simulated unmanned vehicles.

3.0 METHOD

3.1 Participants

Participant recruitment and data collection are underway. Plans are to recruit twenty-four participants (12 males, 12 females) from a U.S. Midwestern university and the U.S. general public. All participants will have normal or corrected-to-normal vision and hearing. Participants are being randomly assigned such that there will be 12 participants for each type of team composition (human-human and human-autonomy). Experimental participants serve as the "operator" in all missions, either delegating tasks to a teammate in the operator-driven team structure or managing preassigned tasks in the role-driven team structure.

3.2 Simulation Testbed

The experimental trials were conducted at the U.S. Air Force Research Laboratory (AFRL) utilizing the IMPACT (Intelligent Multi-UxV Planner with Adaptive Collaborative/Control Technologies) testbed [25] developed by a tri-service U.S. team under the leadership of AFRL. IMPACT (Figure 1) combines several autonomy advancements into a single control station to support the command and control of multiple unmanned vehicles (UVs). These autonomous capabilities include cooperative control algorithms for rapid vehicle route planning [26], intelligent agent reasoning that compares possible courses of actions and makes vehicle allocation recommendation [27], and autonomic monitoring technologies [28].

Figure 1: IMPACT Control Station Simulation Testbed (early version).

Figure 2 illustrates the four monitors used in the participant and teammate control stations. The Tactical Situation Display (TSD) (top monitor) provided a map and current UV information such as pertinent mission locations, UVs and their associated routes, as well as a vehicle panel showing a summary of each UV's status. The bottom monitor provided a 'sandbox' display, which mirrored the TSD but also allowed the operator to create 'what-if' scenarios by generating and comparing possible UV plans before implementing them. The sandbox was also where most of the interfaces that supported operator UV management were located. The right monitor provided a detailed dashboard for each of the twelve UVs that included status information and that UV's sensor feed. Three Dell Precision R7610 workstations running Microsoft Windows 10 located in a different room provided the sensor feeds for the twelve UVs (four feeds per workstation). SubrScene, an in-house simulation visualization toolkit, was used to produce the imagery in the sensor feeds. The left monitor contained a help window and Task Log (to be described later).

Figure 2: IMPACT Testbed Display Layout Employed.

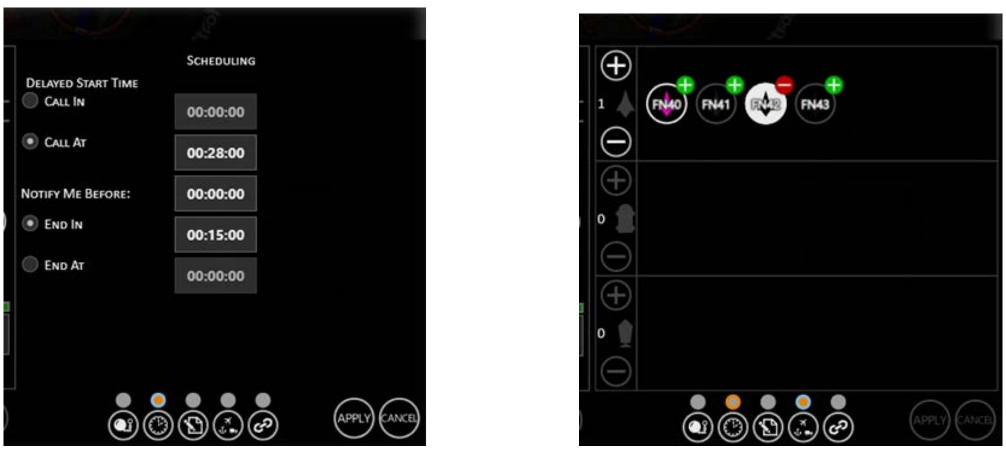
IMPACT was designed to support operator management of multiple heterogeneous UVs (air, ground and sea) in support of base defence missions. During "normal" operations (i.e., no immediate threat identified), there were two types of tasks. One type termed "RAMs" (random anti-terror measures) involved positioning one or more UVs at certain locations at either a specified time or within a time frame. For example, one of the four types of RAMs involved having an air UV circle a particular building every 10 minutes. "Normal base defence events" (NBDEs) were the other task type during normal operations and included checking on unidentified vehicle/building alarms, providing escort (ground/sea UV) or overwatch (air UV) to friendly forces, and getting a UV's "eyes on" to obtain a specified sensor view.

The simulation also supported tasks performed when an intruder was detected until an "all clear" signal was received. Each task type in response to an intruder had specified subtasks to be performed, depending on the event (e.g., perimeter breach, gate runner, mortar fire, crowd forming, and an improvised explosive device detonation). During intruder events, execution of RAMs was discontinued.

A fourth type of task involved responding to a UV failure (e.g., engine or sensor) or an environmental event (e.g., smoke or fog at a location). Finally, the chat interface in IMPACT supported the posing of query tasks, some addressing vehicle state (e.g., speed, altitude, and which vehicle was closest to a certain location) and others designed to assess the participant's situation awareness related to tasks (e.g., "What was the last task completed?"). The chat interface was also utilized for participants to rate their current workload level every 5 minutes of the scenario without pausing the simulation. The participants responded to prompts in the chat window using the Bedford Workload Scale [30].

IMPACT's flexible delegation interfaces supported high level task delegation similar to the interactions that occur in human teams. Base defence tasks were completed using the control station's interfaces that employed a playbook-based approach [31]. One or more UVs could be quickly tasked by selecting the appropriate pictorial coded icon representing a high-level play. Play selection then initiated a series of automatic actions to complete the task (e.g., call a "point inspect" play and specify the location to task a UV to automatically position an on-board sensor at that location). Participants primarily called plays from the Task Manager (see Figure 6), but plays could also be called from a play calling tile or directly from a location or vehicle on the map via a radial menu (these other interfaces are described in [32]-[34]). Once a play was called, IMPACT's intelligent agents [35] determined all the other parameters to complete the task (i.e., which vehicle(s) and route(s)). If desired, the operator could instead define one or more of these parameters as well as other details via the Play Workbook and associated pages (Figures 3 and 4). For example, by selecting the "cloud" icon to the right of the "sun" icon, the participant communicated to the intelligent agent that there was an environmental change that could influence the recommended UV/sensor type. Parameters that could be specified included environmental conditions, optimizations (e.g., time to get to target, tracking capabilities, etc.; Figure 3), a start time or end time for the play (Figure 4a), and UVs to consider for allocation (Figure 4b). Once the participant's inputs were received and the intelligent agent's plan was accepted by the participant, the play icon and details were added to an Active Play Manager (Figure 5a), and the vehicle(s) began executing the selected play.

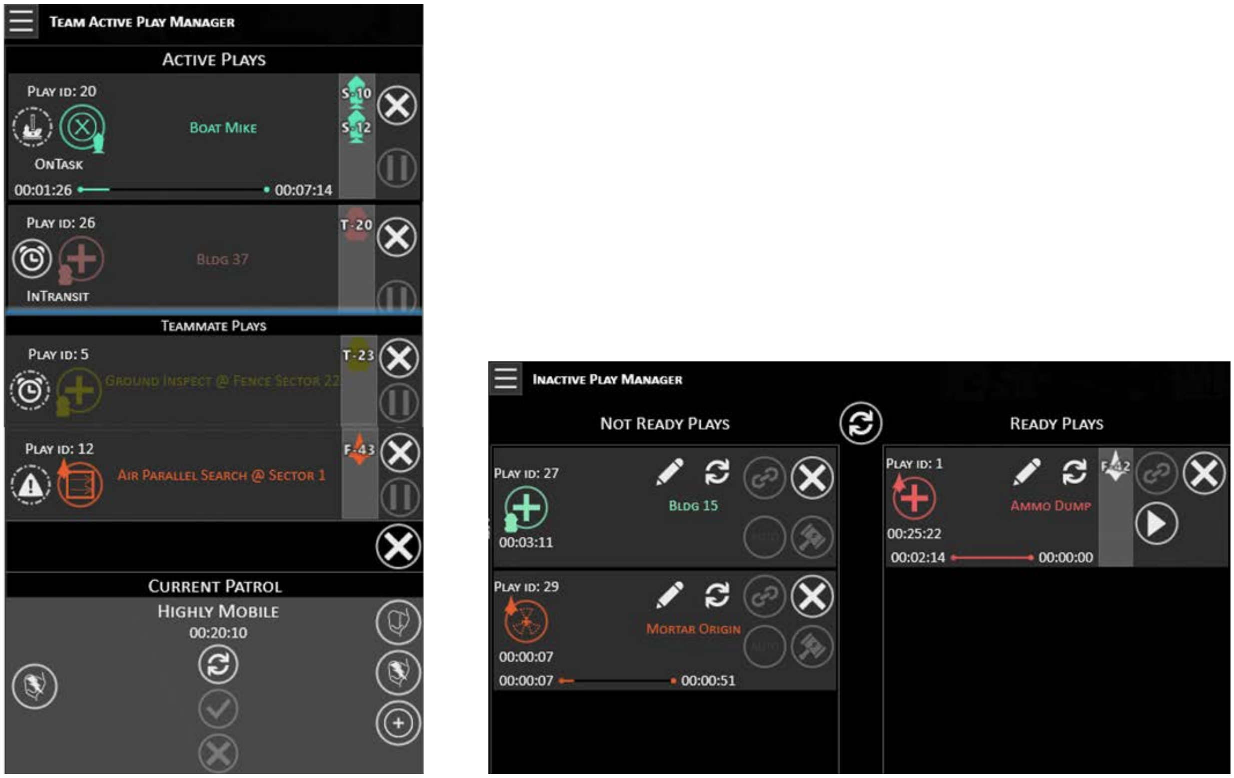
Figure 3: Play Workbook.



a) Scheduling Page

b) Vehicle Selection Page

Figure 4: Play Workbook Pages Accessible via Buttons on the Bottom of the Workbook’s Right Pane.



a) Active Play Manager **b) Inactive Play Manager**

Figure 5: Play Manager Tiles.

3.3 Simulation Testbed Enhancement to Support Teaming

Although the original IMPACT simulation [25] can be viewed as representing the state-of-the art integration of several autonomous capabilities to support an operator's task delegation (via plays), it was still limited with respect to the goal of human-autonomy peer-to-peer type communication and coordination. First, with the delegation scheme, the human operator initiated each play in response to a mission event (the autonomy did not initiate plays in response to a tasking until commanded by the operator). There was no give-and-take between the operator and autonomy in terms of trading tasks or subtasks based on the teammate's capabilities or current situational demands. Even though the autonomy was quite capable for some subtasks, IMPACT's original control and display interfaces failed to provide adequate mechanisms to support operator-autonomy task coordination. Thus, before experimentation began, the IMPACT simulation was enhanced to better support this coordination and enable the evaluation of alternative team structures as a function of mission complexity.

Refinement of IMPACT's interfaces focused on improving the method by which the participant received and responded to tasks, including the ability to coordinate with a teammate. When a mission related event occurred, the participant was notified not only through the chat messaging system, but also with a pictorial icon corresponding to the task that appeared in the left column of a refined Task Manager (Figure 6). Tasks were organized in rows based on task type, with the top row being tasks generally considered highest priority (intruder events) and the bottom row being lowest priority (queries). The line coding used for the circle around the task icon designated if the task was assigned to be completed by the participant (solid), by the teammate (dashed), or involved subtasks with both the participant and teammate sharing the workload (dashed-dotted line). The Task Manager's right pane listed the subtask or steps for completing the task that was selected on the left pane. For example, in Figure 6 the fence alarm task is highlighted in the left pane and its associated subtasks ("send UGV to fence location" and "UAV expanding square search at fence location") are shown in the right pane. The majority of tasks could be completed by a play call (or series of play calls) or, in the case of queries, by sending a chat message. In the human-autonomy teams, the "autonomous teammate" completed tasks/subtasks assigned to it (either because of the initial assignment in the role-driven team structure or by the participant selecting the lightning bolt icon on either the task or a subtask). The same task assignment procedure was used in the human-human teams.

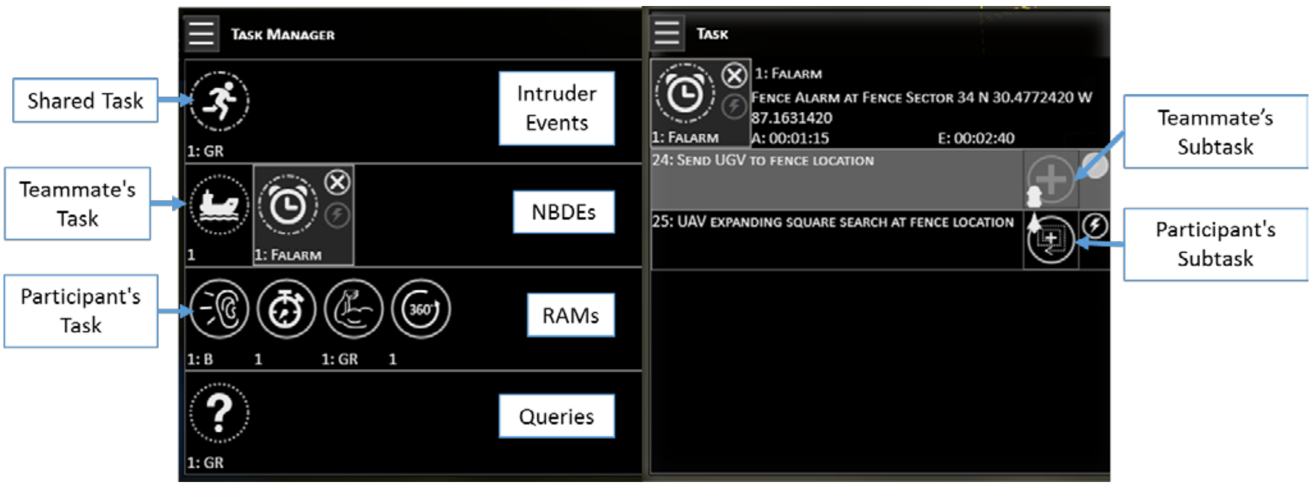


Figure 6: Task Manager.

The participant could remove a task from the Task Manager at any time. After being deleted from the Task Manager, the task was added to a new interface called the Task Log (Figure 7) that maintained a record of all the tasks received as well as details about each task (e.g., when the task was created in the Task Manager and if the operator ("owner"), teammate ("viewer"), or both ("mixed") completed the task. A task could also be sent back to the Task Manager from the Event Log.

Figure 7: Task Log.

The Active Play Manager (Figure 5a) was also revised to include all ongoing plays, grouping the plays called by the participant versus the teammate. Ongoing plays, including those called by the teammate, were able to be cancelled from the participant's Active Play Manager. Any plays that were created but not yet activated (e.g., scheduled to start at a later time or vehicles currently not available) were stored in an Inactive Play Manager (Figure 5b).

These revised interfaces complemented the other IMPACT interfaces to better support human-autonomy teaming and enable the evaluation of operator-driven versus role-driven team structures. The interfaces informed team cognition and promoted communication and coordination between the participant and teammate. This was accomplished by providing more status information about what tasks had been completed and what tasks needed to be completed, as well as providing information about intentions (who was responsible for completing a task, how they would complete it and the constraints to be taken into consideration). The participants also learned about the capabilities of the teammate during training (see 3.4.2 and 3.5.1).

3.4 Experimental Design

The mixed experimental design is illustrated in Figure 8. The team composition (human-human or human-autonomy team) was a between-subjects variable, with each of the 24 participants randomly assigned to a composition resulting in 12 human-human teams and 12 human-autonomy teams. The teammate in both conditions, either human or autonomy, had the same abilities in the experiment. Team structure and mission complexity were varied within-subjects. Each participant completed two half-hour trials, one low complexity trial and then one high complexity trial, with each of the team structures (operator-driven and role-driven), for a total of four trials with the assigned team composition (human or autonomy teammate). Team structure trials were counterbalanced to control for order effects.

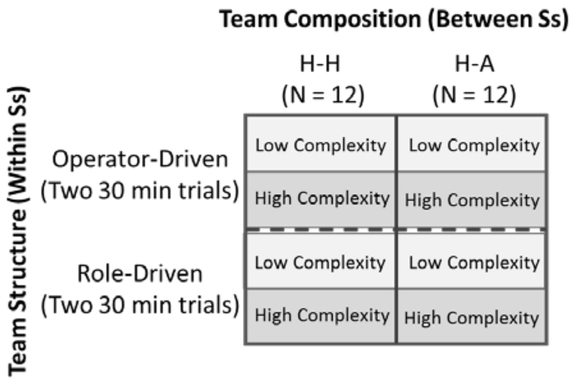


Figure 8: Experimental Design.

Table 1 shows how the levels of mission complexity differed in the number of each task type that needed to be completed. The low and high complexity missions also differed in terms of the interrelatedness and simultaneousness of tasks (and time between tasks). Besides the base defence tasks, participants were also tasked to respond to Commander queries and complete imagery tasks by confirming with a sensor operator through chat that a target was in sight. (The sensor operator role was performed by an experimenter sitting at the Test Operator Console, to be described later.)

Two experimental scenarios were developed for each complexity level (four scenarios in total). For each complexity level, the number and types of tasks (e.g., RAMs, intruder events, vehicle failures, etc.) and the task coordination demands (e.g., number of events that could be completed independently versus jointly) were the same in both scenarios. The scenarios were counterbalanced across the team structure condition, such that the low complexity scenario experienced with one team structure differed from the low complexity scenario in the other team structure; the same procedure was used for the high complexity scenarios.

Table 1: Number of Each Task Type for Low and High Complexity Missions.

	Task Response to Event Involved a:						
	Play(s)			Response	Query		
	RAMs: Random …Measures	**Normal Base Defence**	**Intruder Detected**	**UV or Environ Problem**	**UV Info**	**Task SA**	**Workload Rating**
TOTAL	**8**	**7**	**5**	**5**	**10**	**10**	**12**
Low Complexity Mission	3	2	2	2	5	5	6
High Complexity Mission	5	5	3	3	5	5	6

In order to better evaluate human-autonomy teaming, the mission scenarios were designed to provide situations that required teammate interaction for optimal performance. These included:

- Task coordination – Mission tasks varied on level of coordination, with some events providing opportunities for the participant and teammate to work together to complete a single task and other tasks capable of being completely accomplished by the participant or the teammate.
- Asymmetrical information – Either the participant or teammate obtained information that the other was not aware of (e.g. vehicle failures, environmental changes). The changes in the situation required coordination across the team for successful task completion.
- Demands exceed teammate's capability – Due to certain constraints, the teammate could not complete a task that it normally would be able to. An example was a teammate not being able to respond to an event because there were no available vehicles capable of completing the task and the teammate was not authorized to cancel on going plays to free up vehicles capable of completing the task.

3.4.1 Team Composition

For participants assigned to a human-autonomy team, the additional intelligent agent capabilities were referred to as the "autonomous teammate". The autonomous teammate role was filled by a human confederate manning a duplicate IMPACT control station. (Participants were not informed that a confederate was fulfilling this role.) The autonomous teammate had the capability to complete certain tasks without the need for participant input, such as responding to queries and calling plays in response to an intruder event (see Table 2). The autonomous teammate's capabilities were consistent across the two team structures with the only difference being whether the task assignment was operator-driven or role-driven. The operator and autonomous teammate were responsible for the same mission area and controlled the same number and types of assets (i.e., two air, two ground, and two sea UVs each). However, the teammates could not task each other's assets or trade control of assets. Asset allocation by the autonomous teammate (i.e., play calls) was reflected on the experimental participant's control station display and vice-versa, however the human operator was not able to visually view the displays of the other station. There were limited messages that participants could send through chat to the autonomous teammate. If a chat was sent that was not on that list or not in the proper format, a chat response of "I cannot understand" was sent back to the participant.

The human-human team composition consisted of the experimental participant (operator) and a human teammate that worked together to complete mission goals. The human teammate was played by a confederate whose system actions were constrained such that they matched that of the autonomous teammate's capabilities (see Table 2; participants were told their human teammate was "not authorized" to complete certain tasks rather than "not capable"). The participant and human teammate each used an individual control station to complete shared mission goals. Consistent with the human-autonomy team composition, each teammate was responsible for the same mission area and had control of half the assets. Asset allocation was reflected on both stations. Human teammates were able to communicate freely through chat; they were not limited to the chat message set used in the human-autonomy team.

Besides the participant's control station and the teammate's control station (manned by the confederate acting as the human teammate for the human-human team and as the autonomous teammate for the human-autonomy team), an experimenter-manned Test Operator Console (TOC) was also used. The TOC displays mirrored the sandbox map of the participant's station with the addition of a display that was used to drive the scenario events. Any UV allocation performed by the team members was reflected on the TOC's displays (two Perceptive Pixel Inc. 27 in. LCD monitors @2560 x 1440 resolution).

3.4.2 Team Structure

In the operator-driven team structure, the participant operator received all taskings directly from chat and decided which tasks to complete versus which tasks to assign to the teammate (either human or autonomous). All tasks were populated in the Task Manager as the participant's responsibility but the participant was also able to use the Task Manager to assign tasks or subtasks to the teammate by selecting the lightning bolt icon to the right of the task icon (see Figure 6). Participants were thoroughly trained on which task types the teammate could accomplish. There was also a help menu (left monitor) that could be referenced during the trials.

The role-driven team structure automatically assigned tasks to a teammate based on predefined roles, such that the teammate could complete tasks without the participant's assignment. Participants were trained that a task would only be assigned to the autonomous or human teammate if it was within the scope of its capabilities or authorization and if that task had been predetermined to be the responsibility of that teammate (high level roles for the operator and teammate are shown in Table 2). All tasks were populated in the Task Manager, regardless of the ownership, allowing the participant to be aware of what tasks were assigned to the teammate.

Table 2: Task Assignments for the Role-Driven Team Structure.

Participant	Teammate
• Subset of NBDEs • RAMs • Queries related to task completion	• Subset of NBDEs • Intruder Events • Queries related to vehicle status or allocation

For both team structures, participants changed task allocation between their teammate and themselves using the Task Manger (new tasks) or the Active Play Manager and chat (ongoing tasks). The teammates utilized chat messages and the Task Manager to communicate limitations or constraints to the participant (e.g., chat messages that tasks were outside their capabilities or authority and sending tasks back to the participant in the Task Manager along with a chat message as to why tasks could not be completed, such as "no available assets").

3.5 Procedure

Each participant took approximately eight hours to complete the experiment. Training and two of the experimental trials were conducted in the morning with one team structure, followed by the third and fourth experimental trials in the afternoon with the other team structure. Upon arrival, participants were given a brief overview of the study and then read and signed an informed consent document. A background questionnaire was also administered asking age, gender, vision, hearing, and video gaming experience. Lastly, participants were asked to fill out the Desirability for Control Scale [36], Propensity to Trust Machines Questionnaire [37], and the Ten-Item Personality Inventory [38].

3.5.1 Training

Initial training consisted of the participants being given a detailed overview of the IMPACT system layout, displays, and control functionality. Training covered all the abilities of the simulation testbed, as well as its limitations. Following formal training, participants were provided a review followed by approximately 30 minutes for them to practice interacting with the system by responding to a series of mission events representative of what they would encounter during the experimental trials. This also provided an opportunity for participants to gain experience working with the teammate, to better understand the teammate's limitations and help develop a shared mental model. The experimenter was nearby monitoring each participant's actions in order to answer any questions about the system (such as why something happened or how to perform a certain action) or to intervene if necessary (for instance, if a situation occurred that a participant could not recover from, the experimenter provided an explanation and restarted the practice trial).

Subsequent training focused on how to interact and work with the assigned (human or autonomous) teammate. Participants were trained on which tasks their teammate was capable/authorized to complete, how to trade tasks between team members, and how to use chat to communicate with their teammate. This training consisted of procedures that were common to both team structures. Training on the specifics related to the two different team structures was given prior to the start of the corresponding experimental trials with that team structure (e.g., what tasks would be automatically assigned to their teammate for the role-driven team structure). An approximately 30-minute capstone scenario was completed before the experimental trials for each team structure to assess whether the participant had been trained to a sufficient level of performance. The capstone was self-paced and included every type of task they would experience during an experimental trial.

3.5.2 Experimental Trials

Participants completed four 30-minute trials, two with each of the two team structures (operator-driven and role-driven). For each team structure, the first 30-minute trial was at the low mission complexity level and the second 30-minute trial was at the high mission complexity level (see Table 1).

The following task performance data were collected: number of tasks completed, time to complete the tasks, and the correctness of tasks completed. Time to complete a task was measured as the time from when the task was received in chat to when all subtasks had been completed. Correctness of task completion was represented by an experimenter calculated summary score based on four subcomponents: (1) Did the participant select the correct location/target? (2) Did the participant select the correct play? (3) Did the participant choose the optimal vehicle or sensor? and (4) Did the participant meet the event's constraints? Accuracy and response times were collected for the participants' responses to system failures, environmental events, and commander queries. The number of tasks completed by each teammate (or jointly) and the number of tasks that were traded between the teammates, along with any associated task details (e.g., task type, task name, subtasks, etc.), were also collected. In addition to all system interactions (e.g., chats sent, buttons clicked, etc.), voice audio and video of the participant's sandbox display were recorded for all trials.

Subjective data were also recorded. A questionnaire was administered after trials were completed with each team structure. Its questions addressed the system's usability (SUS; [39]) and participants' situational awareness and trust (Checklist for Trust; [40]) for the trials just completed. At the completion of all experimental trials, a final questionnaire was given with items to compare the two team structures. An experimenter also observed participants during the scenarios and took note of any incidents of automation surprise or moments of confusion; these were later discussed in the post-mission debrief that followed questionnaire completion.

4.0 RESULTS / SUMMARY

Data collection and analyses are underway; reported results should be viewed as preliminary. Complete results will be documented in a follow-on publication. In this section, questionnaire results are first described with respect to participants' workload ratings, situation awareness assessments, and perceived performance on mission tasks. These data are examined for each team structure: operator-driven and role-driven. Next, initial data pertaining to characteristics of the teamwork between the experimental participant and teammate (autonomous and human) are presented. This includes assessment of the team processes, and the degree of communication and coordination. Finally, example comments and questionnaire ratings pertaining to the collaborative interfaces employed to support teaming are provided.

4.1 Participants' Workload, Situation Awareness, and Perceived Performance

Participants' ratings indicated that workload was higher for the operator-driven team structure compared to the role-driven team structure. This result likely reflects the additional workload associated with managing task assignment in the operator-driven structure as compared to the role driven structure where tasks were automatically assigned to the teammate. Other contributing factors to the higher workload ratings for the operator-driven structure include the requirement for the participant to be aware of available assets for both team members (as resources were limited), as well as their teammate's capabilities, in order to appropriately assign tasks. Moreover, the participants' workload was impacted whenever the teammate sent tasks back to the participant. Some participants even commented that they preferred to complete as many tasks as possible themselves, until they ran out of assets, forcing them to assign tasks to their teammate. In contrast, participants' ratings indicated less workload associated with the role-driven team structure.

These initial results on workload differences between the two team structures are similar to previously reported results examining workload differences between management-by-consent and management-by-exception automation control schemes [41]. The management-by-consent scheme (operator required to make a consent response before autonomy completes the task) is similar to the operator-driven team structure used in the present experiment in that both required the operator to engage in additional decision making and associated control station inputs for task completion. In contrast, with the role-driven team structure and the management-by-exception scheme, the operator has less steps to complete: initial task assignment is automated in role-driven teams and tasks are automatically completed in management-by-exception schemes unless the operator decides to intervene. The similarities in workload ratings between a role-driven team structure and management-by-consent are noteworthy since participants in the present experiment only had to make high-level task/play commands and the simulation's autonomy capabilities handled the workload involved in vehicle allocation and route planning.

The results pertaining to the situation awareness ratings are also of interest. Past research has shown that operators often have decreased situation awareness when task completion is highly automated (i.e., the higher the automation level, the more the operator is "out of loop"). However, initial data in the present study show that despite task assignment being automated in the role-drive team structure, the participants rated their situation awareness higher compared to the operator-driven structure where they were more involved in task assignment. This can be accredited partially to the role-driven structure having a pre-established working agreement prior to the start of missions. Therefore, participants were already aware of what tasks would be automatically assigned to their teammate and those that would be assigned to themselves. Additionally, not having the extra workload managing task assignment provided participants spare capacity to monitor control station displays and this may have increased awareness of the current state of each UV and task completion. The role-driven team structure was also rated better in terms of both perceived individual and team performance. Moreover, participants' frequency in shedding the low priority, but demanding RAM tasks was less when the team structure was role-driven, compared to operator-driven, perhaps also reflecting the spare capacity available without the management overhead of making task assignments. Both team structures, though, showed that shedding RAM tasks was more frequent in the high complexity missions, compared to the low complexity missions.

4.2 Teamwork: Communication and Coordination in Task Sharing

Experimenter observations and task sharing data indicated that when participants became overwhelmed in the high complexity missions, they more frequently assigned tasks to their teammate even when the tasks could not be completed by the teammate. This resulted in the teammate sending those tasks back to the participants, increasing participants' workload even more. This strategy to rely on the teammate in the operator-driven team structure to send tasks back, rather than exert workload to judge whether the task should even be assigned to the teammate, likely reflects participants' inability to maintain awareness of the teammate's capabilities and asset availability during the high complexity missions. This result is also aligned with the lower ratings for the operator-driven team structure, compared to role-driven, on questionnaire items addressing collaborative climate.

The rate of chat communications was slightly higher when the teammate was human compared to being autonomous. However, even in the human-human team configuration, chat communications were fairly limited. These results may reflect features of the training, design of the control station, and participants' tasks. Participants were trained on teammate capabilities and tasks that could be handled by the teammate prior to the experimental trials, eliminating the need for chat communications to obtain this information. Also, the symbology used on the displays was designed to facilitate quick retrieval of status information (e.g., available assets and task assignment) and the Task Manager play-based interface facilitated efficient task assignment. Thus, control station design may have made it more likely that needed information could be efficiently acquired from the displays rather than by taking the time to employ the chat system. Future research will examine the frequency and content of participant-teammate communications with the control station modified to also accommodate voice communications for information exchange. The use of missions with more dynamic, unexpected changes may also prompt more frequent communications to better evaluate the adequacy of communication interfaces.

4.3 Interface Feedback

The refined Task Manager was found to be useful in both team structures regardless of whether the teammate was autonomous or human. Questionnaire comments indicated that the overall look and feel was effective as well as the interface's functionality in supporting task management (e.g., knowing the tasks assigned to each team member and whether they have been completed or not). However, initial questionnaire ratings suggest that the Task Manager supported coordination better for the role-driven team structure compared to the operator-driven structure in terms of understanding the teammate's actions. These ratings may also reflect that the teammate's actions were more defined in the role-driven structure.

Despite the positive ratings for the Task Manager interface, participants had comments on how it could be improved. Some feedback indicated a desire for a more salient indication of task priority, perhaps adding color-coding to the task icons in addition to the row locations used to differentiate task priority. Participants' feedback also indicated that having more task status information would be useful in the Task Manager and Task Log. Desired improvements include task progression status and a mechanism whereby all chat messages/information related to a specific task, beyond the initial chat that created the task (e.g., updates, all clears, etc.), are easily accessed.

5.0 FUTURE RESEARCH

Mixed-initiative systems allowing for autonomy to share in task completion for reasons such as workload management are not inherently good or bad. For the benefits of a mixed-initiative team to be realized, however, the system's design must ensure human-autonomy coordination breakdowns are mitigated and tools are provided that promote effective teamwork [42]. Thus, interfaces [21] need to be refined in parallel to advancements in the capabilities of autonomy, especially with the vision of a human operator and intelligent agent dynamically collaborating to problem solve and share task completion in a manner similar to effective human teams.

Advancement of intelligent agent technologies is also needed to support reliable operation, provide better autonomy transparency, and support human-autonomy collaboration [5]. The fact that the present experiment had to employ a human confederate to serve as the autonomous teammate illustrates the need for more capable intelligent agent support. Moreover, improvements are essential to enable more dynamic task distribution and trading to help equalize workload across the team and capitalize on the capabilities of each team member. Enhancements are also required such that the autonomy exhibits stable performance in a range of operating conditions and mission events, notifying the human teammate when it is operating outside of its competency bounds, just as an operator would do in a team of humans.

Future development of technology that better supports peer-to-peer type communication between human-autonomy team members is also necessary. For instance, identification of effective interface functionality and bi-directional communication technologies is needed to better support task negotiations between human and autonomy team members. This would allow human-autonomy teams to create working agreements pertaining to task sharing and determine under what conditions task swapping should be entertained or even automated. Design efforts to improve collaboration between a human operator and an autonomous teammate should result in interfaces that inform efforts that are addressing the vision of a well-situated team of multiple humans working together with multiple autonomous systems.

6.0 REFERENCES

[1] Baker, D.P., Day, R. and Salas, E. (2006). Teamwork as an essential component of high-reliability organizations. *Health Services Research*, *41*, 1576-1598.

[2] Salas, E., Sims, D.E., and Burke, S.C. (2005). Is there a ‘big five’ in teamwork? *Small group research*,*36*, 555-599.

[3] Bradshaw, J.M., Feltovich, P.J., Johnson, M.J., Bunch, L., Breedy, M.R., Eskridge, T. and Uszok, A. (2008). Coordination in human-agent-robot teamwork. *2008 International Symposium on Collaborative Technologies and Systems*, 467-476.

[4] Christoffersen, K. and Woods, D.D. (2002). How to make automated systems team players. *Advances in Human Performance and Cognitive Engineering Research*, 1-12.

[5] Johnson, M., Bradshaw, J.M., Hoffman, R.R., Feltovich, P.J. and Woods, D.D. (2014). Seven cardinal virtues of human-machine teamwork: Examples from the DARPA robotic challenge. *IEEE Intelligent Systems*, *29*(6), 74-80.

[6] Klein, G., Woods, D.D., Bradshaw, J.M., Hoffman, R.R. and Feltovich, P.J. (2004). Ten challenges for making automation a “team player" in joint human-agent activity. *IEEE Intelligent Systems*, *19*, 91-95.

[7] DeChurch, L.A. and Mesmer-Magnus, R. (2010). The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of Applied Psychology*, *95*(1), 32-53.

[8] Wilson, K.A., Salas, E., Priest, H.A. and Andrews, D. (2007). Errors in the heat of battle: Taking a closer

look at shared cognition breakdowns through teamwork. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(2), 243-256.

[9] Urban, J.M., Bowers, C.A., Monday, S.D. and Morgan Jr., B.B. (1995). Workload, team structure, and communication in team performance. *Military Psychology*, *7*, 123-139.

[10] Cannon-Bowers, J.A. and Salas, E. (2001). Reflections on team cognition. *Journal of Organizational Behavior*, *22*, 195-202.

[11] Salas, E. E. and Fiore, S.M. (2004). Team cognition: Understanding the factors that drive process and performance. *American Psychological Association*.

[12] Mathieu, J.E., Heffner, T.S., Goodwin, G.F., Salas, E. and Cannon-Bowers, J.A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, *85*(2), 273.

[13] Atkinson, D.J., Clancey, W.J. and Clark, M.H. (2014). Shared awareness, autonomy and trust in human-robot teamwork. *Artificial Intelligence and Human-Computer Interaction (AIAA) Spring Symposium Series*.

[14] Ososky, S., Schuster, D., Jentsch, F., Fiore, S., Shumaker, R., Lebiere, C. and Stentz, A. (2012). The importance of shared mental models and shared situation awareness for transforming robots from tools to teammates. *International Society for Optics and Photonics, SPIE Defense, Security, and Sensing, Vol. 8387*.

[15] Espinosa, A., Kraut, R., Slaughter, S., Lerch, J., Herbsleb, J. and Mockus, A. (2002). Shared mental models, familiarity, and coordination: A multi-method study of distributed software teams. *ICIS 2002 Proceedings*, 39.

[16] Jonker, C.M., Van Riemsdijk, M.B. and Vermeulen, B. (2011). Shared mental models. *Coordination, Organizations, Institutions, and Norms in Agent Systems*, *6*, 132-151.

[17] Kleinman, D.L. and Serfaty, D. (1989). Team performance assessment in distributed decision making. *Proceedings of the Symposium on Interactive Networked Simulation for Training*, 22-27.

[18] Michan, S. and Rodger, S. (2000). Characteristics of effective teams: A literature review. *Australian Health Review*, *23*(3), 201-208.

[19] Rothwell, C. and Shalin, V. (2017). Human-Machine articulation work: Functional dependency dialogue for human-machine teaming. *19th International Symposium on Aviation Psychology*, 270.

[20] Harbers, M., Jonker, C. and Van Riemsdijk, B. (2012). Enhancing team performance through effective communication. *Proceedings of the 4th Annual Human-Agent-Robot Teamwork Workshop*, 1-2.

[21] Kilgore, R. and Voshell, M. (2014). Increasing the transparency of unmanned systems: Applications of ecological interface design. *International Conference on Virtual, Augmented and Mixed Reality*, 378-389.

[22] Urban, J.M., Weaver, J.L., Bowers, C.A. and Rhodenizer, L. (1996). Effects of workload and structure on team processes and performance: Implications for complex team decision making. *Human Factors: The*

Journal of the Human Factors and Ergonomics Society, *38*(2), 300-310.

[23] Bowers, C.A., Urban, J.M. and Morgan Jr., B.B. (1992). The study of crew coordination and performance in hierarchical team decision making. Technical Report No. TR-92-01. Orlando: University of Central Florida.

[24] La Porte, T. and Consolini, P. (1998). Theoretical and operational challenges of "high-reliability organizations": Air-traffic control and aircraft carriers. *International Journal of Public Administration*, *21*, 847-852.

[25] Draper, M., Rowe, A., Douglass, S., Calhoun, G., Spriggs, S., Kingston, K. … Frost, E. (2018). Realizing autonomy via intelligent adaptive hybrid control: Adaptable autonomy for achieving UxV RSTA team decision superiority (also known as Intelligent Multi-UxV Planner with Adaptive Collaborative Control Technologies (IMPACT). (2018). Final Report. AFRL Technical Report (AFRL-RH-WP-TR-2018-0005). Wright-Patterson Air Force Base, OH.

[26] Kingston, D., Rasmussen, S. and Mears, M. (2009). Base defense using a task assignment framework. *AIAA Guidance, Navigation, and Control Conference*, 6209.

[27] Douglass, S. (2013). Learner models in the large-scale cognitive modelling (LSCM) initiative. In R. Sottilare, A. Graesser, X. Hu and H. Holden (Eds.), *Design Recommendations for Intelligent Tutoring Systems, Volume 1*, 111-126. Orlando, FL: U.S. Army Research Laboratory.

[28] Dzieciuch, D., Reeder, J., Gutzwiller, R.S., Gustafson, E., Coronado, B., Martinez, L., Croft, B. and Lange, D.S. (2017). Amplifying human ability through autonomics and machine learning in IMPACT. *Proceedings Society of Photo-Optical Instrumentation Engineers (SPIE), Micro- and Nanotechnology Sensors, Systems, and Applications IX, Volume 10194*, p. 101941Y. International Society for Optics and Photonics.

[29] Draper, M., Calhoun, G., Hansen, M., Douglass, S., Spriggs, S., Patzek, M., Rowe, A. …Frost, E. (2017). Intelligent multi-UxV planner with adaptive collaborative/control technologies (IMPACT). *Proceedings of the International Symposium on Aviation Psychology*. Dayton OH: Wright State University. https://isap.wright.edu/isap-pubs.

[30] Roscoe, A.H. and Ellis, G. A. (1990). A subjective rating scale for assessing pilot workload in flight: A decade of practical use. Technical Report No. RAE-TR-90019. United Kingdom: Royal Aerospace Establishment.

[31] Miller, C.A. and Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(1), 57-75.

[32] Mersch, E.M., Behymer, K.J., Calhoun, G.L., Ruff, H.A. and Dewey, J.S. (2016). Game-based delegation interface design for unmanned vehicles: Color coding and icon row assignment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*, 122-126.

[33] Calhoun, G.L., Ruff, H.A., Behymer, K.J. and Mersch, E.M. (2017). Operator-autonomy teaming interfaces to support multi-unmanned vehicle missions. In P. Savage-Knepshield and J. Chen (Eds.),

Advances in Human Factors in Robots and Unmanned Systems: Advances in Intelligent Systems and Computing, 499, 113-126. Switzerland: Springer.

[34] Calhoun, G., Ruff, H., Behymer, K. and Frost, E. (2018). Human-autonomy teaming interface design considerations for multi-unmanned vehicle control. *Theoretical Issues in Ergonomics Science, 19*, 321-352.

[35] Hansen, M., Calhoun, G., Douglass, S. and Evans, D. (2016). Courses of action display for multi-unmanned vehicle control: A multi-disciplinary approach. *Artificial Intelligence and Human-Robot Interaction (AIAA) Fall Symposium Series*.

[36] Burger, J.M. and Cooper, H.M. (1979). The desirability of control. *Motivation and Emotion, 3*(4), 381-393.

[37] Merritt, S.M. and Ilgen, D.R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors, 50*(2), 194-210.

[38] Gosling, S.D., Rentfrow, P.J. and Swann, W.B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504-528.

[39] Brooke, J. (1996). SUS: a 'quick and dirty' usability scale. In P. Jordan, B. Thomas, I. McClelland, and B. Weerdmeester (Eds.), *Usability Evaluation in Industry*. Bristol, PA: Taylor and Francis Inc.

[40] Jian, J.Y., Bisantz, A.M. and Drury, C.G. (1998). Towards an empirically determined scale of trust in computerized systems: Distinguishing concepts and types of trust. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 42*(5), 501-505.

[41] Ruff, H.A., Narayanan, S. and Draper, M.H. (2002). Human interaction with levels of automation and decision-aid fidelity in the supervisory control of multiple simulated unmanned air vehicles. *Presence: Teleoperators & Virtual Environments, 11*(4), 335-351.

[42] Borst, C., Flach, J.M. and Ellerbroek, J. (2015). Beyond ecological interface design: Lessons from concerns and misconceptions. *IEEE Transactions on Human-Machine Systems, 45*(2), 164-175.