2022

User Interface Design for Supervisory Control of Multiple Manned and Unmanned Air Vehicles

Taleri Lynn Hammack
Wright State University

USER INTERFACE DESIGN FOR SUPERVISORY CONTROL OF MULTIPLE MANNED AND UNMANNED AIR VEHICLES

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

By

TALERI LYNN HAMMACK

M.S., Wright State University, 2015

B.S., University of Idaho, 2012

2022

Wright State University

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

FEBRUARY 24, 2022

I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION BY Taleri Lynn Hammack ENTITLED User Interface Design for Supervisory Control of Multiple Manned and Unmanned Air Vehicles BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy.

John Flach, Ph.D.
Dissertation Co-Director

Ion Juvina, Ph.D.
Dissertation Co-Director

Scott N. J. Watamaniuk, Ph.D.
Graduate Program Director

Debra Steele-Johnson, Ph.D.
Chair, Department of Psychology

Barry Milligan, Ph.D.
Vice Provost for Academic Affairs
Dean of the Graduate School

Final Examination

Kevin B. Bennett, Ph.D.

Kyle J. Behymer, Ph.D

The views expressed in this dissertation are those of the author and do not necessarily reflect the official policy or position of the Army, the Department of Defense, or the United States Government.

The Tasking and Execution of Collaborative Unmanned and Manned Systems with Autonomy (TECUMSA) technologies described within this dissertation were developed through research funded by the United States Army Combat Capabilities Development Command (DEVCOM) Aviation & Missile Center (AvMC) Aviation Development Directorate (ADD). An in-depth description of a cognitive work analysis for this project is published as a technical report (FCDD-AMT-20-12).

ABSTRACT

Hammack, Taleri Lynn. Ph.D., Department of Psychology, Wright State University, 2022. User Interface Design for Supervisory Control of Multiple Manned and Unmanned Air Vehicles.

This dissertation research will cover lessons learned from the three-year, iterative design and evaluation of TECUMSA (Tasking and Execution of Collaborative Unmanned and Manned Systems with Autonomy). TECUMSA is a graphical user interface and autonomous tool suite that enables a single operator (e.g., an Air Mission Commander) to team with autonomous capabilities (e.g., route planning, aircraft task allocation) to effectively command and control multiple manned and unmanned aircraft in a contested battlespace. The user/AMC was responsible for accomplishing a series of reconnaissance, surveillance, and threat neutralization tasks in a hostile and dynamic simulated battlespace. The main challenges in this problem space are cognitive bandwidth of operators (e.g., maintaining situation awareness, allocating attention flexibly across multiple aircraft), and their ability to coordinate and collaborate with subordinate autonomous agents. The main objective of this research was therefore determining what control mechanisms offered the TECUMSA operator stability and reliability of control. Two formal system evaluations will be discussed, where a total of 15 Army aviators used TECUMSA to complete multiple hours of simulated air assault operations in a synthetic task environment. This research explored distributed supervisory control, where the operator distributed authority to automation for continuous manual control tasks using

Play Calling (i.e., directability). The following research will also cover observations from the system evaluations highlighting interface features that afforded the user the ability to observe, perceive, and understand the state of the world relative to their goals and intentions (i.e., observability). One of the major themes in this dissertation is the importance of observability and directability as design principles, and the implications they have for both user interface design and human-autonomy teaming.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

To my advisor Dr.John Flach, there will never be enough thank you's to properly express my gratitude. You have shared a lifetime of wisdom throughout my graduate career. You have always been incredibly gracious with your time and knowledge, and made time to advise my dissertation even after formally retiring from teaching. Your guidance and insights kept me moving forward, I always left our meetings feeling re-inspired to write. Thank you for helping me finish the academic journey we started together 9 years ago.

To Dr.Brian Dyre and the graduate students in his Visual Performance Laboratory at the University of Idaho (U of I), for giving me my first human factors research opportunity exploring display design using optic flow information in a flight simulator.

To Dr.Todd Thorsteinson, my undergraduate advisor at U of I for proofreading all my applications to graduate schools and encouraging me to pursue a doctorate program.

To Infoscitex Corp. and the U.S. Army DEVCOM – AvMC, for allowing me to use TECUMSA as part of my dissertation. A special thanks to Dr.Thomas Alicia, SUMIT's program lead, and to Nobletech Solutions Inc. for providing invaluable aviator expertise throughout the SUMIT effort.

Additional thanks to my committee members Dr.Kevin Bennett and Dr.Ion Juvina, who have served on both my thesis as well as my dissertation committee. It has been a pleasure collaborating with you both, thank you for sharing so much of your time and expertise. A special thanks to Dr.Kyle Behymer for not only serving on my dissertation committee but for bringing me on to the Infoscitex team to work on the SUMIT program; thank you for all of your support and mentorship in my professional career.

To my wife, Chase, thank you for being by my side for 11 years, including during the countless hours of homework, before every nerve-racking presentation, throughout every late night of studying, and when I came home one day ranting about this incredible thing called "human factors." You were there when we had to move 2000 miles away from home so I could attend graduate school. You are how I know home was never a place but a person. Thank you for being my rock during this dissertation and every day before and since. You have always given me encouragement, patience, wisdom, and most importantly happiness. I am so excited to continue loving and growing with you.

To my dad, thank you for being my first teacher. You have always fostered my passion for learning and problem-solving. Thank you for encouraging me to ask questions and for giving me the freedom to learn by doing.

To my grandma, Bev, thank you for always encouraging and supporting me, and reminding me to take time when I need it. You always drove me to be resilient and empowered. It means the world to me to make you proud, and I am certain I wouldn't be where I am now without you. -With all my love.

DEDICATION

To my Grandma & Wife

CHAPTER 1 - Theoretical Motivation

Perception-Action Coupling

James J. Gibson (1966, 1979) proposed an ecological approach to perception whereby perception is a product of a person's interaction with their environment. The theory posits that as one moves through the world they are continually gathering new sources of optical sensory information. Gibson emphasized that this movement throughout the world is essential as it creates optical flow where the ecological optics either change or are invariant but in either case will offer the perceiver additional information.

Gibson's theory opened the door to a paradigm shift away from there only being extrinsic metrics of perception (i.e., observer independent) that are only concerned with physical qualities of the sensory input, to there also being *intrinsic* metrics that are observer dependent. So for example, size relative to a meter stick is extrinsic, whereas size relative to the user's hand is intrinsic. These intrinsic metrics of perception correspond to the *functional* properties of the objects of interest, or *action possibilities* (e.g., seeing size in terms of "can you grab the object") that Gibson termed "affordances." He writes: "An affordance cuts across the dichotomy of subjective-objective and helps us to understand its inadequacy. It is equally a fact of the environment and a fact of behavior. It is both physical and psychical, yet neither. An affordance points both ways, to the environment and to the observer" (p. 129).

Action-specific perception is therefore rooted in Gibson's theory, as it argues that people perceive their environment in terms of their ability to act within it. This means that as actors (i.e., people who take action, not Hollywood celebrities), we are not only moving through physical space but through functional space as well. In other words, action possibilities are not merely specified by the extrinsic properties of the object but also the perception of action potentials as the actor moves and acts in their environment.

There are numerous examples of action-specified perception. For example, softball players who performed better perceived the ball as bigger (Witt & Proffitt, 2005), tennis players who successfully returned the ball perceived it as slower-moving (Witt & Sugovic, 2010), and people wearing heavy backpacks judged objects as being farther away than people who were not encumbered by a heavy backpack (Proffitt, Stefanucci, Banton, & Epstein, 2003). The key point these findings illustrate is that affordances are defined not in the action, and not in the perception, *but rather in the coupling of the perception and action*. Affordances specify both what you *can* do (action-driven) and what you *should* do (perception-driven).

Consider walking into your kitchen for a cup of coffee and being asked to determine the diameter of your cup. This size judgement (i.e., the diameter) is grounded in physical space. However, because we are acting in functional spaces, the cup's perceived size is specified relative to the possibilities for action. So rather than needing to discriminate the cup's size based on changes in millimeters, you just need to know the size at which the cup changes from graspable to not graspable. Graspability in this case is a functional, affordance-based judgement.

Perception and action are not separable processes, and their coupling has significant implications for user interface design. Information requirements of a user shift as the context of their activity shift. This means that system operators have to be able to see the world scaled to her/his action capabilities. In other words, the user interface must make the relevant possibilities and constraints *observable* to the operator (e.g., incongruences between actual- and desired-system state, ability to predict future system states), as well as *directable* (e.g., ability to make inputs to achieve desired system state).

These design principles of *observability* and *directability* are critical aspects of system design (e.g., running a nuclear power plant, flying an attack helicopter), but particularly challenging for graphical user interface design since the operator is no longer able to *directly* "move" and "act" in their environment. Instead, the ecology must be represented in a way that is both perceivable and actionable via the interface, so that the user can evaluate what they can/want to do (perception), do it (action), and compare what happened to what they intended (coupling outcome with intent). The interface must be designed for information processing in a closed-loop feedback system where, *as the user interacts with the world (via the interface), so too should they be continually sampling information from it.*

The key tenant of the cyclical interdependence between perception and action is accepting that the perceived *consequences* of behavior will continually shape the *prioritization* of behavior (i.e., beliefs, intent, expectations), which guides future actions and their subsequent consequences. Consider a commercial pilot's goals/intent while flying an aircraft. The highest level goal of the pilot is to not fatally crash the plane. To achieve this goal they require *observability* of the field of safe travel and the ability to *direct* how the field of safe travel is accomplished (e.g., changing heading, coordinating with ground control to maneuver around a weather system). When possibilities and constraints of action are specified relative to the goals of the actor (e.g., displaying a stall risk if airspeed drops below a certain value) then you can generally expect the operators to make good choices (e.g., increase airspeed).

The quality of performance is ultimately determined by the quality of the coupling between perception and action, as illustrated in Figure 1. The functional space actors move through is defined by their beliefs, intentions, and expectations, as well as the consequences of their behavior. Additionally, there are design factors that can impact the quality of perception (e.g., supporting situation awareness, adjusting visual saliency and clutter) as well as the quality of action (e.g., adjust ease of interact-ability, function allocation), as indicated by the green boxes in Figure 1. The orange box in Figure 1 suggests some of the behaviors that contribute to the quality of team coordination, where numerous perception-action cycles must appropriately align to accomplish the situational goals (e.g., training and skill development, increasing common understandings between team members) (Cooke et al., 2013; Klein, Wiggins & Dominguez, 2010; Larue et al., 2020).

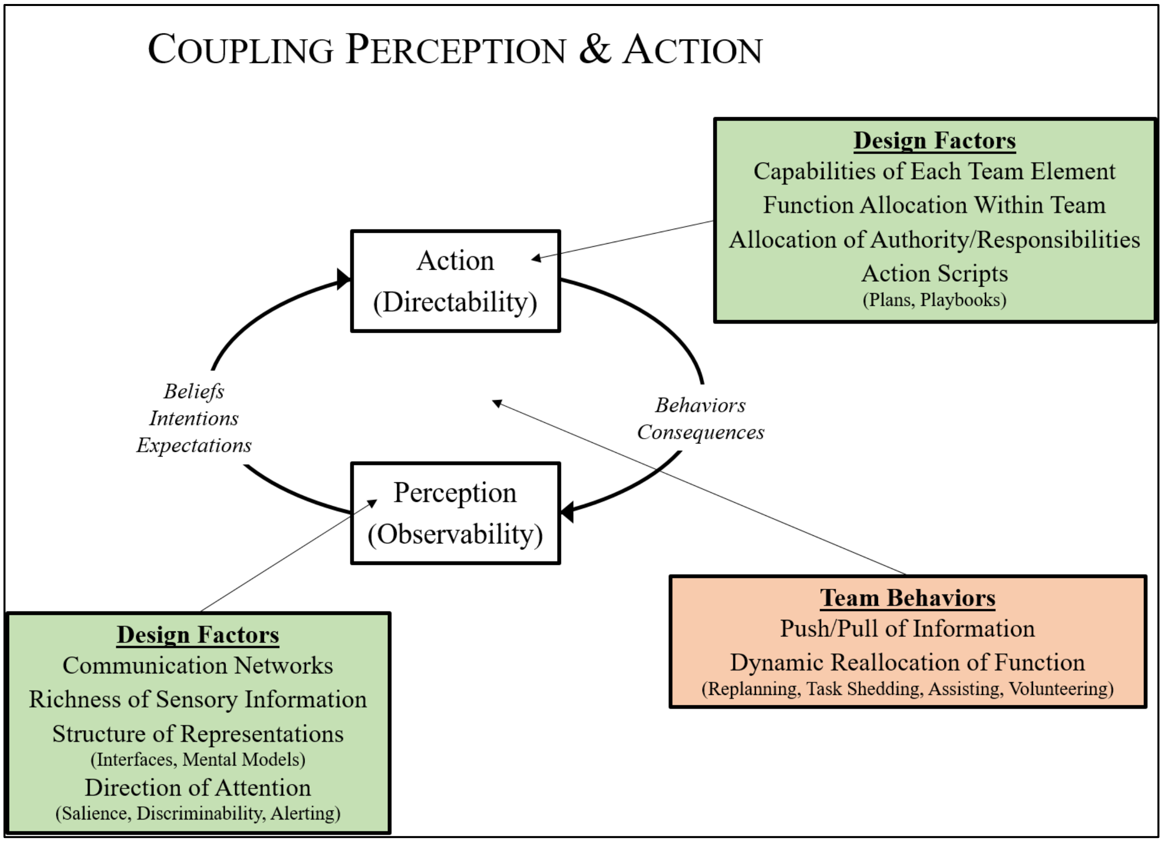


Figure 1. Coupling Perception & Action.

Ultimately, the demands of the situation determine what affordances are relevant. Operators do not need to be made aware of all of the available affordances in a system, but rather the *affordances that are contextually relevant* to their mission, activities, and goals. Referencing back to the discussed affordances of a coffee cup, designers must keep in mind that they do not have to “display” all of the graspable sizes. As such, operator decision-making may best be served by simply representing the *boundary conditions* where an object moves from graspable to not graspable, or where it moves from graspable with *one* hand to graspable with *two* hands (Flach, 1988). *Transitional states* are a good example since they are functional boundary conditions with *critical shifts in action possibilities* for an operator that therefore must be adequately specified or represented.

These transitional states will likely occur when context (e.g., system capabilities, environmental conditions), intent, or consequences change. So, as the perception-action cycle updates, the critical shift in action possibilities and consequences must be appropriately specified relative to the operator's goals. Consider when my goal shifts from needing a vessel to drink from, to a vessel to pot a plant in. In this case, holes at the bottom make the object unsatisfactory for my thirst needs and yet highly desirable for my gardening needs.

The design of objects, or interfaces, needs to make the possibilities and consequences for action apparent so that the elicited response is appropriate relative to the user's goals. If a ceramic flowering pot with holes in the bottom is given a handle the size of four fingers, the design is likely to fail since the unsuspecting user is likely to discover the holes only after they end up with a lap full of coffee. Or take for example an airplane transitioning from a stable flight into an impending stall. In this case, the critical action boundaries have shifted significantly; even a slight upward pitch of the aircraft's nose is now highly consequential, and worse yet the proper inputs to mitigate the stall will not be clear if the interface (and operator's mental model) do not appropriately update to the new set of action constraints.

It is also essential to represent the higher-order relationships among those variables that specify relevant action boundaries. However, there can be more than one right way to provide pertinent information to a user. Consider a problem where distance and time to traverse between objects is important to avoid accidentally flying into the firing range of an enemy threat. The first step of the design is therefore to determine *what* information is pertinent in defining a safe path of travel (e.g., avoiding mountainous terrain, areas within enemy firing range). Only after determining what needs to be represented can the interface designer determine *how* to represent the critical action boundaries. As cognitive systems engineers, we must first identify that this time-space relationship is a critical action boundary, as it specifies the possibilities and constraints on action. For example, it is not possible to reach the threat in less than five minutes (a *possibility for action*) without the aircraft's path of travel crossing in range of an enemy threat that will likely shoot it down (a *consequence of action*).

Observability can only be specified relative to directability, and vice versa. In other words, the possibilities for action are determined by *both* what is observable and what is directable. For example, maintaining a safe braking distance in your vehicle is not possible without considering your breaking capacity, or *directability* (e.g., are the breaks worn, are the roads icy). However, you must also consider perceptual feedback such as the changing distance between you and the objects or vehicles in front of you (e.g., changing information to the sensorimotor system, retinal flow), or *observability*. As a driver, determining you will not be able to stop before the next intersection going 45 miles per hour on an icy road is the result of coupling your perceptual information and action feedback, particularly if you have no experience driving on icy roads. Hence the importance of a rich coupling between observability-directability in interface design.

Summary

This dissertation explores the path of developing a graphical user interface to enable a user to not only utilize advanced autonomous technologies under normal operational conditions but also allow them to cope and adapt to the unanticipated complexities that are inevitable in a dynamic and uncertain world. In line with the ecological interface design approach (Vicente & Rasmussen, 1992; Bennett & Flach, 1992; Pawlak & Vicente, 1996; Bennett, Posey, & Shattuck, 2008), this goal requires that the user is able to understand the current state of the complex sociotechnical system (i.e., the work domain, ecology) and glean insights to predict future states via the graphical user interface (*observability*). Additionally, the user must be able to take the necessary actions within the user interface to accomplish the dynamic goals of their work, including the distribution of responsibilities and coordination of goals with an autonomous teammate (*directability*). The following chapters will therefore cover human-autonomy coordination and the possibility for coordination failure, synchronization of goals & priorities, and the distribution of authority.

CHAPTER 2 - Domain, Evaluations, and Timeline Narratives

TECUMSA Overview

This dissertation explores TECUMSA[1] (Tasking and Execution of Collaborative Unmanned and Manned Systems with Autonomy), a graphical user interface and autonomous tool suite that enables a single operator (e.g., an Air Mission Commander) to team with autonomous capabilities (e.g., route planning, aircraft task allocation). A major goal of TECUMSA's design was to explore solutions and improvements for autonomy and decision aiding technology that would enable a single operator, sitting in a helicopter out in a contested battlespace, to effectively command and control multiple manned and unmanned aircraft to complete an air assault mission.

One of the prominent design approaches for accomplishing supervisory control of a team of unmanned vehicles is to use the "Play Calling" metaphor (Miller, Goldman, Funk, Wu, & Pate, 2004; Fern & Shively, 2009; Apker, Johnson, & Humphrey 2016; Calhoun et al. 2013). The play calling approach allows the human operator to issue high-level commands and objectives, while the autonomy handles the low-level details of route and waypoint planning and aircraft tasking recommendations (for a more detailed discussion of supervisory control see Chapter 7). This allocation of responsibilities utilizes the strengths of both humans and machines (Cummings, 2014). As such, TECUMSA was designed so that the human was the high-level decider of goals and

[1] TECUMSA leverages work previously done for IMPACT (Intelligent Multi-UV Planner with Adaptive Collaborative/Control Technologies), a system developed for military base defense operations utilizing multiple heterogeneous unmanned systems (for more details see Draper et al., 2018).

objectives (tasking aircraft through "play calling"), and the autonomy's role was deciding the most efficient way to accomplish the task at hand.

Method

Design and Development Cycles

For this project, two formal design and development cycles took place. The *first cycle* was an 11-month design and development phase followed by a formal evaluation of the system, hereon referred to as the "Round 1 Evaluation." The *second cycle* of design and development was a 10-month phase followed by a second formal evaluation of the system, hereon referred to as the "Round 2 Evaluation." Figure 2 illustrates the timeline of the project, showcasing the iterative design process with subject matter experts (SMEs) and the formal system evaluations in FORCE (Future Open Rotorcraft Cockpit Simulation Environment), an aircraft cockpit with battlespace simulation software that will be described in the following section.

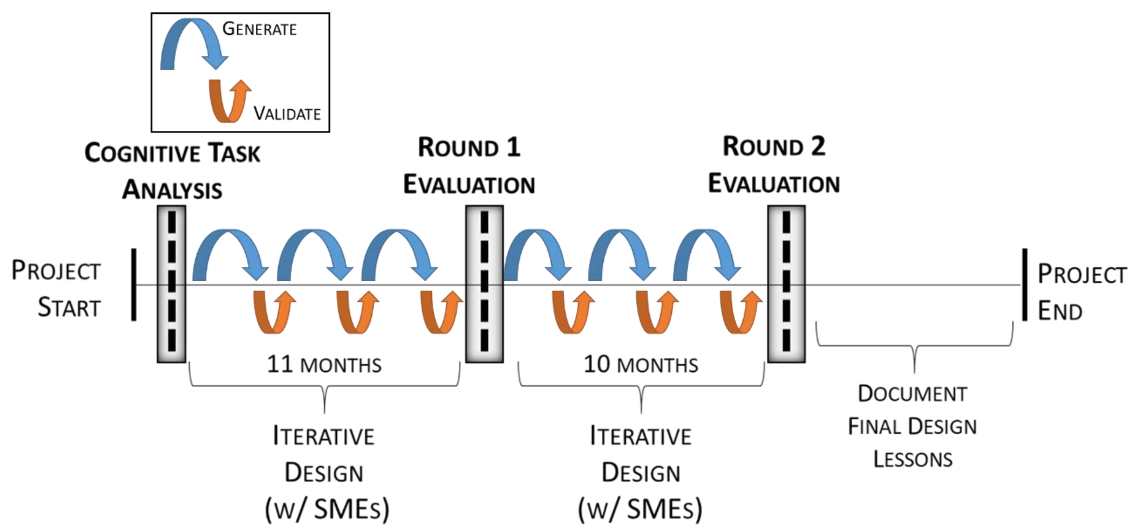


Figure 2. Iterative Design and Evaluation Cycle.

FORCE Simulator

In both the Round 1 and Round 2 Evaluations of TECUMSA the FORCE simulator was used. FORCE is a Government-owned and operated simulation environment containing a stationary medium-fidelity cockpit (see Figure 3). The FORCE simulator provided physical realism for the individuals that participated in the evaluation. Operational fidelity in FORCE was achieved through the use of the Virtual BattleSpace (VBS) terrain simulation, containing vegetated and mountainous terrain, infrastructures (e.g., roads, buildings), mobile entities including enemy threats and neutral civilians, damage models for living and inanimate objects, flight models of the manned and unmanned aircraft (e.g., limits on airspeed, bank angles), and a variety of aircraft payloads (e.g., missiles, electro-optical imaging (EO), laser designators).

It is worth noting that each participant could see a very different sequence of events within the same vignette due to the highly fidelity entity modeling in VBS. For example, because the people and vehicles actually moved throughout the battlespace and were reactive to events (e.g., ran away if an explosion occurred nearby), one participant could have seen hostile vehicles in the landing zone, while another participant may never see them (e.g., hostile vehicles may have already been intercepted by the participant earlier in the vignette and thus never arrive). The responsiveness of entities in the battlespace thus provided a valuable opportunity to observe how participants responded to a wide variety of events.

The ultimate goal of this research was to design a system that enabled a single operator – an Air Mission Commander (AMC), paired with a suite of autonomous capabilities to manage a diverse team of manned and unmanned aircraft during an air assault mission. To achieve this goal, research questions had to be grounded in properties of real-world Army Aviation air assault mission environments. Continuous collaborations with Army Aviation and Army Artillery SMEs helped to derive a laboratory environment that was representative of the operational conditions in real-world missions. This is also known as a *synthetic task environment,* since meaningful abstractions were pulled from the real-world domain to make a realistic environment with relevant operational tasks (e.g., aircraft capabilities and limitations, dynamic enemy behavior, time sensitive decision-making) (Cooke, Rivera, Shope, & Caukwell, 1999). Note, the key of a synthetic task environment is not necessarily the physical fidelity of the simulator, but how well the research questions are mapped to the semiotics of the actual work domain.

Figure 3. Picture of Soldiers using the Future Open Rotorcraft Cockpit Environment, or FORCE, Simulator (Bradley, 2019).

Equipment

In both the Round 1 and Round 2 formal evaluations, TECUMSA was integrated into the FORCE simulator. As seen in Figure 3, the FORCE simulator was configured with two side-by-side, forward-facing crew station positions. The participant was stationed in the left seat of the cockpit, which was equipped with two mounted multi-touch touchscreen displays, a trackball mouse cursor, and a QWERTY keyboard for both rounds of evaluations. In the right seat was an experimenter serving as the manned aircraft pilot. The FORCE simulator included a virtual out-the-window visual of the battlespace (e.g., terrain, infrastructures).

General Air Assault Mission

In both the Round 1 and Round 2 Evaluations each participant performed four simulated air assault missions, hereon referred to as *vignettes*. All vignettes focused on two of the primary stages in an air assault operation, based on Army Field Manual No. 3-99 (2015, March): (1) “Movement to Landing Zone” and (2) “Landing Zone Preparation.” The general situation in these vignettes is that the AMC is issued a hasty air assault mission (i.e., minimal time for pre-mission planning), and by the end of the thirty-to-sixty-minute vignette will have needed to accomplish the following mission objectives:

- *Primary Objective*: Search the *air ingress route* for enemy threats & clear this route of hostile activity,
- *Primary Objective*: Search the *landing zone* (LZ) for any threats or obstacles & clear the area in and around the LZ of any hostile activity,

- *Secondary Objective*: Search designated named areas of interest (NAIs) for enemy activity.

Pop-up events were also scheduled for all of the vignettes used in the Round 1 and Round 2 Evaluations. The injected pop-up events were pre-scripted events or information transmissions, which were not known ahead of time by the participant, and intended to add complexity to the vignette and potentially change their originally planned course of action as an AMC. An example from Round 1 of a pop-up event was the scheduled loss of one of the AMC's unmanned aircraft due to enemy fire. An example of a pop-up event from Round 2 was a hostile enemy ambush of friendly Ground Troops, where the troops request immediate assistance from the AMC's team as they are being attacked.

These pop-up events were not intended to take permanent precedence over the AMC's Primary Objectives in the mission. Often, however, pop-up events would take priority over the Secondary Objectives since most of the pop-up events involved some form of enemy activity that could impact the Primary Objectives. Figure 4 shows a reproduced example of the map layout for vignette 4 in the Round 2 Evaluation, including *primary objectives* (high priority) in teal, *secondary objectives* in green, *pop-up event objectives* (high priority) in orange, and *No Fly Areas* (NFAs) in red. All objectives required at least one aircraft to visually inspect and/or provided armed support.

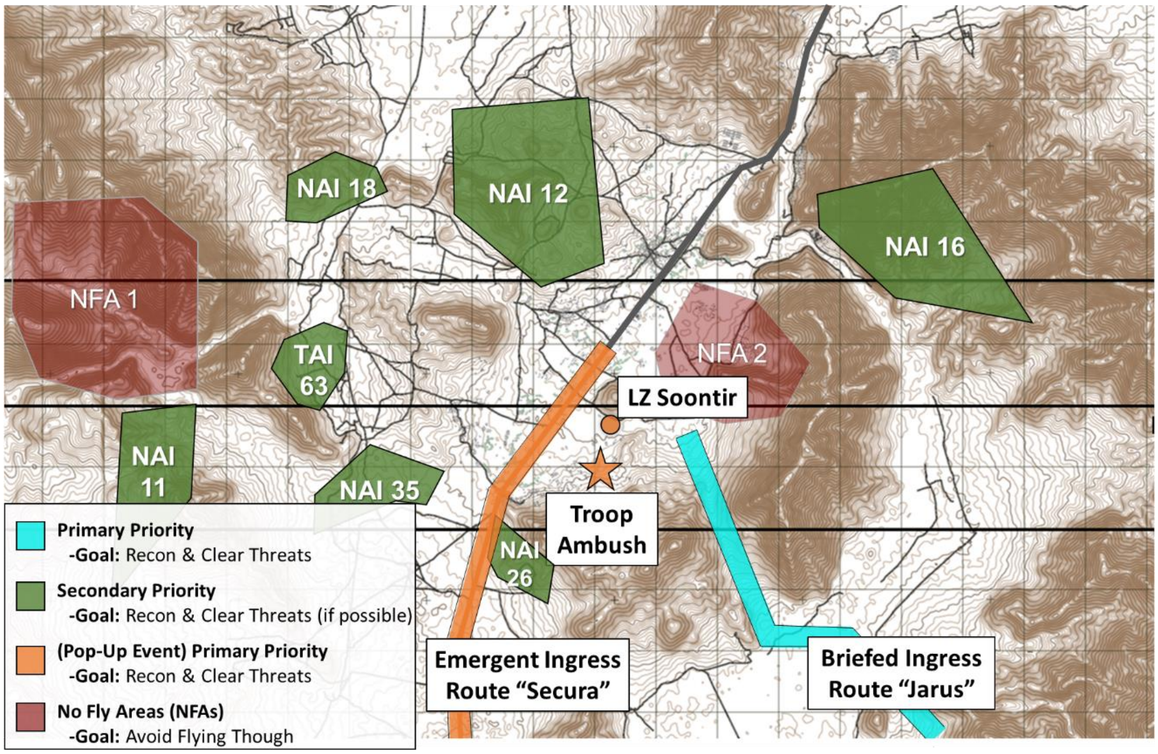


Figure 4. Reproduced example of map layout in Vignette 4 of the Round 2 Evaluation. (Note, Named Area of Interest (NAI) and Target Area of Interest (TAI)).

Participant Responsibilities as Air Mission Commander

In both Round 1 and Round 2 Evaluations, the participants acted as the AMC of the air assault mission. The participant used the TECUMSA system to effectively manage their available team of aircraft and resources (e.g., missiles, rockets, camera sensors) to accomplish tasks that are common in Army air assault operations. In these evaluations, the AMC was responsible for accomplishing a series of reconnaissance, surveillance, and threat neutralization tasks.

The participants were in the left seat of the FORCE simulator, "riding" in an aircraft modeled similarly to a UH-60 Black Hawk helicopter and referred to as the AMC's "Ownship." To accomplish all of their mission objectives, the participant not only had their own aircraft (i.e., Ownship) but also two to seven unmanned air vehicles

(UAVs) under their command and control. The size of this manned and unmanned aircraft team varied depending on the intended complexity of the vignettes in the Round 1 and Round 2 Evaluations (discussed further in the *Empirical Design* section).

The participant's tasking of the Ownship and supporting UAVs was primarily accomplished via play calling. The TECUMSA system offered three main types of plays, reconnaissance-based plays for finding threats in the battlespace, surveillance-based plays for keeping a camera sensor fixed on an entity, and engagement-based plays for kinetically striking and neutralizing threats. So, if for example, the participant wanted to do reconnaissance of an air ingress route (see Route Jarus in Figure 4), they could simply call an "Air Route Inspect at Route Jarus," after which TECUMSA's autonomous tool suite would recommend an aircraft and plan the quickest flight path to accomplish the play. The participant could then either accept or modify the recommended play before approving it to begin.

The Ownship that the AMC was seated in for the simulation was manned, and thus had a human pilot that the participant could verbally ask to execute a particular flight maneuver instead of, or in addition to tasking the Ownship via play calling. This meant that the participant could verbally tell their Ownship pilot to "fly South of this ridgeline until reaching Phase Line Alpha," or they could call a point inspect play in TECUMSA at the desired location along Phase Line Alpha, which would generate a route that the Ownship pilot could then follow. The Ownship was a member of the experimental team and served as a sort of "human autopilot" who was in charge of waypoint following, but they could not give any strategic input or feedback to the AMC. This meant that the participant remained the sole commander and decision-maker for the manned and unmanned aircraft team's course of action.

As the AMC, participants were ultimately responsible not only for the tasking of their manned and unmanned aircraft team but also for each aircraft's respective payload. There were two main payload control tasks in the Round 1 and Round 2 Evaluations. The first was monitoring and periodical manual steering of the electro-optical (EO) camera sensors, and the second was confirming missile/rocket release. Not all aircraft in the evaluations were equipped with a weapons payload, but all aircraft were equipped with an EO camera sensor to allow for live video feeds of the simulated battlespace. The participant's role in controlling the camera sensor payload typically involved zooming and panning the camera. These adjustments to the live video feed helped the participant to make positive identification (PID) of potential threats, and assess damage after kinetically engaging threats. The participant's role in controlling the weapons payload is discussed in more detail in the *TECUMSA: Human-Autonomy Teaming During Engagements* section.

In Round 1, participants were aided by simulated automatic object *detection* capabilities. This meant that each of the aircraft's video feeds was able to report if a person or vehicle had been detected in the video feed, provided the camera sensor was not zoomed out too far. The TECUMSA system would populate the digital map with an icon for each of the detected people or vehicles, collectively referred to as "entities." Figure 5 and Figure 6 highlight what could be seen in TECUMSA's map display when entities were detected; Figure 5 shows the five vehicle dashboards with live video feeds, and Figure 6 shows the TECUMSA map with entity icons (yellow icons) in and outside of the aircraft's camera sensor field of view (teal projection on the map). Note, a more in-depth discussion of the TECUMSA interface can be found in Chapter 3. Icons in the map were the extent of the visual/auditory feedback for newly detected entities in the Round 1 TECUMSA design.

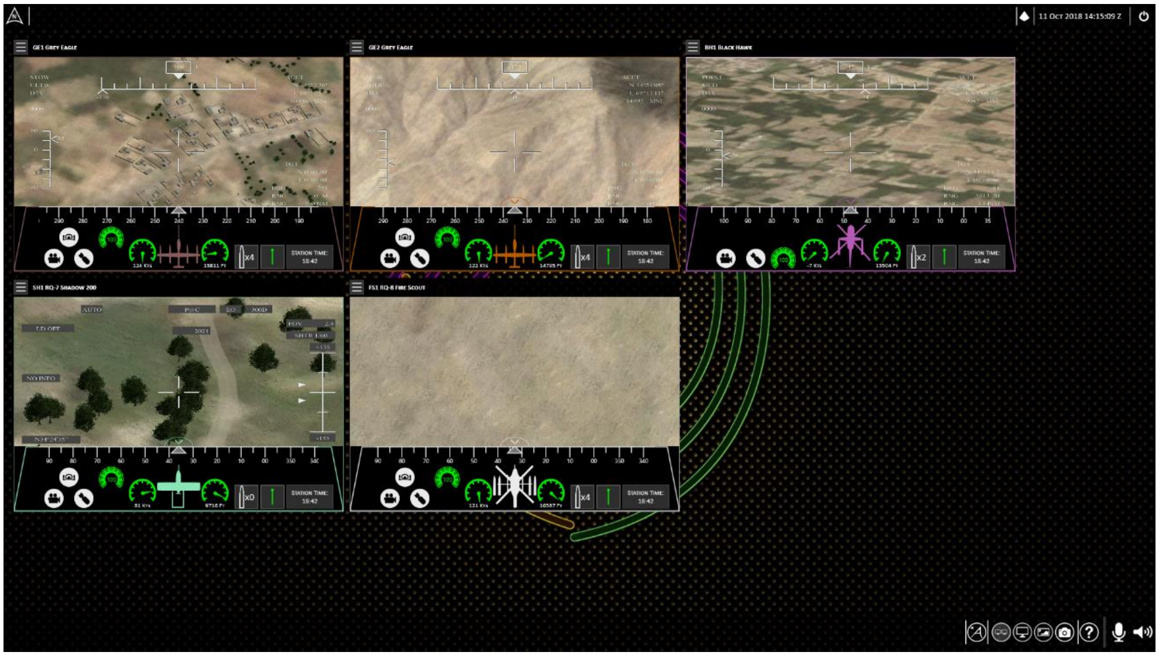


Figure 5. Screenshot of Round 1 TECUMSA Interface, Left Monitor.

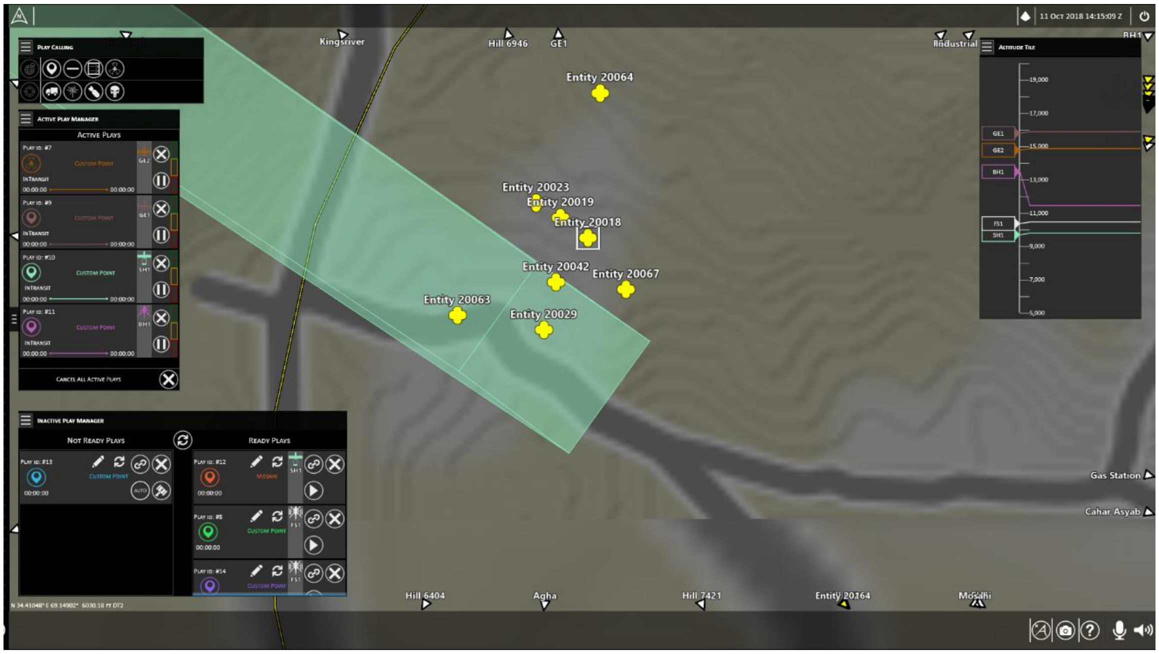


Figure 6. Screenshot of Round 1 TECUMSA Interface, Right Monitor.

In Round 2, the participants were further assisted by simulated automatic object *recognition* capabilities. This meant that each of the aircraft's video feeds was able to report if a person, wheeled vehicle (e.g., pickup truck), tracked vehicle (e.g., tank), or other manufactured objects (e.g., aircraft) had been recognized in the video feeds. Just as in Round 1, when an entity was recognized, the TECUMSA system would populate the participant's digital map with icons for each of the recognized entities. In addition, a list of entities in the aircraft's current camera field of view was shown in a column to the left of each video feed (see Figure 7).

Figure 7. Single Aircraft's Vehicle Dashboard from TECUMSA's Round 2 Interface.

Empirical Design

Participants: Round 1

Although this paper focuses mainly on the Round 2 evaluation, it is worth noting the empirical design of the Round 1 evaluation since TECUMSA was iteratively designed off of the Round 1 results. In the Round 1 evaluation, eight current or former United States Army Pilots with relevant piloting experience (e.g., operational combat, Air Mission Commander, Apache pilot) participated. Participant ages ranged from 26-to-57-years-old (M = 39.4). Additionally, the average number of total flight hours across both fixed-wing and rotary-wing aircraft per participant ranged from 385-to-7200-hours (M = 2641.9). The average number of career time spent as an AMC ranged from 0-to-5000-hours (M = 1018.8). All participants had normal or corrected to normal vision.

Participants: Round 2

In the Round 2 empirical evaluation, seven current United States Army Pilots with relevant piloting experience (e.g., operational combat, Air Mission Commander, Apache pilot, Black Hawk pilot) participated. These pilots were different than the ones in

Round 1. Participant ages ranged from 33-to-45-years-old (M = 37.7). Additionally, the average number of total flight hours across both fixed-wing and rotary-wing aircraft per participant ranged from 295-to-5750-hours (M = 2777.7). The number of career time spent as an AMC ranged from 0-to-2000-hours (M = 918.6). All participants had normal or corrected to normal vision.

Vignette Designs: Round 1 and Round 2

In both the Round 1 and Round 2 Evaluations the four mission vignettes used ranged in complexity. In both rounds of evaluations, the first two vignettes that participants experienced were lower in complexity, and the two latter vignettes were higher in complexity (relative to the prior vignettes). The distinction in vignette complexity was determined through a variety of factors that are listed in Table 1 and Table 2 (Round 1 and Round 2 respectively), including:

- *Complexity*: Relative complexity of each vignette.
- *Mission Duration*: The scheduled duration of the vignette from start to finish.
- *AMC's Total Number of Aircraft*: The AMC's available aircraft for tasking, including the AMC's Ownship.
- *Total Number of Threats*: The total number of threats in the battlespace, including threats to Ground Troops as well as threats to the AMC's aircraft.
- *Number of High Priority Threats*: The total number of threats in the battlespace with Anti-Aircraft Artillery, which are only a threat to the AMC's aircraft. In Vignette 4 of the Round 2 Evaluation, three technical vehicles were also considered high priority because they were actively attacking Ground Troops in the vignette.
- *Total Number of Entities*: The total number of civilian, enemy, and friendly entities on the ground in the battlespace.
- *Number of Named Area of Interest* (NAI) and *Target Area of Interest* (TAI): The number of predefined geographical areas in the battlespace to be reconned when

possible. Note, NAIs and TAIs have distinct information requirements that can be satisfied for the mission.

- *Total Number and Type of Pop-Up Events*: The total number of pop-up events, and the general response to be elicited from the AMC (e.g., 1 pop-up event where the AMC should command at least one aircraft to provide armed overwatch assistance to civilians).
- *Number of AMC's Weapons*: The number of weapons available across all of the AMC's aircraft payloads. Note in addition to missiles, Round 2 also included the use of rockets and Long Range Precision Fires (LRPFs) (see Table 2).

Table 1. Round 1 Experimental Design.

ROUND 1	**VIGNETTE 1**	**VIGNETTE 2**	**VIGNETTE 3**	**VIGNETTE 4**
Complexity	Low	Low	High	High
Mission Duration	30 Minutes	30 Minutes	30 Minutes	30 Minutes
AMC's Total # of Aircraft	3	3	5	5
Total # of Threats	5	3	7	5
# of High Priority Threats (AAA)	1	1	4	2
Total # Entities in AO	75	72	75	97
# NAIs	3	3	4	6
(Total #) Type Pop-up Event(s)	N/A	(1) Provide Armed Overwatch of Civilians	(1) Inspect cause of AMC's Shot Down UAV	(1) Provide Armed Support of Ground Troops
# Missiles	4	4	12	12

Table 2. Round 2 Experimental Design.

ROUND 2	VIGNETTE 1	VIGNETTE 2	VIGNETTE 3	VIGNETTE 4
Complexity	Low	Low	High	High
Mission Duration	30 Minutes	30 Minutes	30 Minutes	60 Minutes
AMC's Total # of Aircraft	7	7	8	8
Total # of Threats	13	22	22	20
# of High Priority Threats (AAA)	4	4	8	10
Total # Entities in AO	61	95	86	117
# NAIs/TAIs	5	4	4	7
(Total #) Type Pop-up Event(s)	(1) Provide Armed Support of Downed Helicopter	(2) Inspect Enemy Activity, Provide Armed Overwatch of Civilians	(3) Inspect Three Separate Reports of Enemy Activity, one being near to LZs	(3) Inspect Enemy Activity, Provide Armed Support of Ground Troops, Provide Armed Support of MEDEVAC
# Missiles / Rockets / LRPF	6 / 9 / ∞	6 / 9 / ∞	10 / 14 / ∞	10 / 14 / ∞

Procedure: Round 1 and Round 2

Each participant evaluation took two days. On the first day participants were given guided trained on TECUMSA followed by two practice runs on a training vignette (note the training vignette was comparable to Vignette 3 in terms of the AMC's number of aircraft, number of high priority threats, and number of NAIs/TAIs). On the first practice run (i.e., "Untimed Training") an experimenter sat next to the participant so they could ask questions and get help, on the second practice run (i.e., "Timed Training") participants completed it without experimenter guidance. On the second day, participants were given a brief refresher in the morning and an opportunity to ask any final questions before starting their four mission vignettes for data collection. Training was not considered completed until the participant had found, identified, and engaged at least one threat.

All participants completed the vignettes in a fixed-order, starting with Vignette 1 and ending with Vignette 4. Although there are benefits to counterbalancing conditions for experimental research (e.g., control the effects of variables), there are costs as well (e.g., increase in number of participants needed per condition). The use of counterbalancing was not compatible with the goals of this research. This is because the vignettes were deliberately designed to increase in complexity throughout the day (e.g., increasing the number of threats, areas of interest, and the aircraft to manage) so experimenters could observe how participants adapted their techniques, strategies, and courses of action.

Throughout the evaluation, a variety of questionnaires were also filled out by participants (see *Appendices* for more details). The following is the general layout of the two-day evaluation:

- Day 1:
    - Informed Consent & Demographics Questionnaires
    - Classroom Training Slides (approx. 20 minutes)
    - Hands-on Training in FORCE (approx. 4 hours)
    - Untimed Training Vignette (approx. 40 minutes)
    - Timed Training Vignette (35-minute time limit)
- Day 2:
    - Refresher Training (averaged approx. 25 minutes)
    - Vignette 1 (30 minutes)
        - Post Vignette Questionnaire
    - Vignette 2 (30 minutes)
        - Post Vignette Questionnaire
    - Vignette 3 (30 minutes)
        - Post Vignette Questionnaire
    - Vignette 4 (30 minutes in Round 1, 60 minutes in Round 2)
        - Post Vignette Questionnaire
        - Post Evaluation Questionnaires

Dependent Measures

A variety of questionnaires were given including the situation awareness rating technique (SART) (Taylor 1990), NASA-Task Load Index (Hart & Staveland, 1988), and the Bedford Workload Scale (Roscoe, 1984). However, the freedom of participants to take different courses of action in a reactive simulation meant that their conversational feedback was more valuable to designers than their scale ratings or scorings. For that reason, the most meaningful data came from the live observations of participant performance, supplemented with feedback on the Post Vignette and Post Evaluation questionnaires (see appendices).

CHAPTER 3 - TECUMSA Interface

TECUMSA: Overall Layout and Tiles

The remainder of this dissertation will primarily focus on the TECUMSA interface design used in the Round 2 Evaluation. The following writing includes an overview of how participants could accomplish their reconnaissance, surveillance, and engagement tasks using TECUMSA, including a description of the overall layout, the play calling process, and human-autonomy task allocation. TECUMSA's *right monitor*/screen (see Figure 8) was dedicated to mission planning and monitoring, where the user could develop plans to accomplish various tasks during the series of air assault mission simulations. The *left monitor* of TECUMSA contained a limited number of vehicle dashboards that could be swapped across aircraft, which presented flight instrument information, payload information, and the live video feed for each of the aircraft's camera sensors. In addition, the left monitor showed an entity list that organized incoming images taken of entities discovered in aircraft video feeds, as well as a section for monitoring high priority areas in the battlespace.

The human-machine interface (HMI) framework for TECUMSA in the Round 2 Evaluation supported touch, mouse and keyboard, gamepad, and game controller inputs. For the default layout, TECUMSA had a map that the operator could pan, rotate, zoom, and tilt from 2D to a 2 ½ D view rendered using Digital Terrain Elevation Data (DTED). The map contained graphics including aircraft icons and air routes, points of interest, areas of interest, NFAs, as well as primary roads and highways. Additional windows, referred to as “tiles,” were overlaid either on the map or on a second monitor. These tiles can be seen in the screenshot of the TECUMSA interface (see Figure 8), which included:

Included in the right monitor (see the right monitor in Figure 8) was the *“Play Calling & Management” tiles*, a series of tiles for managing and monitoring the execution of mission tasks, including:

- Play Calling tile – Contained a variety of plays to execute mission tasks.
- Active Play Manager tile – Contained all ongoing plays.
- Inactive Play Manager tile – Contained all paused or queued plays.

Also included in the right monitor was an *“Altitude Awareness” tile*, which contained the current and projected altitude of all aircraft under the user’s command and control. A *“Digital Map,”* which contained a map of the battlespace, including representations for aircraft, routes, loiter patterns, and pre-defined areas such as landing zones and points of interest. Additionally, icons appeared in the map where known people or vehicles (i.e., entities) were located, which required detection through an aircraft’s camera sensor to trigger simulated automatic target recognition (ATR). A *“Chat” tile*, which allowed users to exchange information via text-based communications. In the Round 2 TECUMSA system, the Chat tile contained two main types of information. The first type was messages between the participant and the simulated tactical operations center (TOC), who was played by an experimenter. The second type was notifications, which included gunfire detection alerts, enemy missile launch detection alerts, and LRPF status alerts.

Included in the left monitor (see the left monitor in Figure 8) was an *“Entity List,”* which contained a list of images taken for entities detected in the battlespace through the aircraft camera video feeds. The Entity List also contained play calling and entity reclassification options. A *"Priority Area Monitoring" tile* which contained recent images taken of the user-designated high priority area(s) of the battlespace (e.g., landing zone alpha) as well as configurable audio alerts for those areas. *"Vehicle Dashboard" tiles*, which contained a live video feed from an aircraft's camera sensor, as well as flight information from the selected aircraft including airspeed, heading, and payload. Vehicle Dashboards also contained play calling and entity reclassification options. Note the *"(Swappable) Vehicle Dashboard"* seen in Figure 8 allowed the user to change which unmanned aircraft's vehicle dashboard was viewed, whereas the *"(Ownship) Vehicle Dashboard"* contained the vehicle dashboard specifically for the Ownship aircraft (i.e., fixed on Ownship).

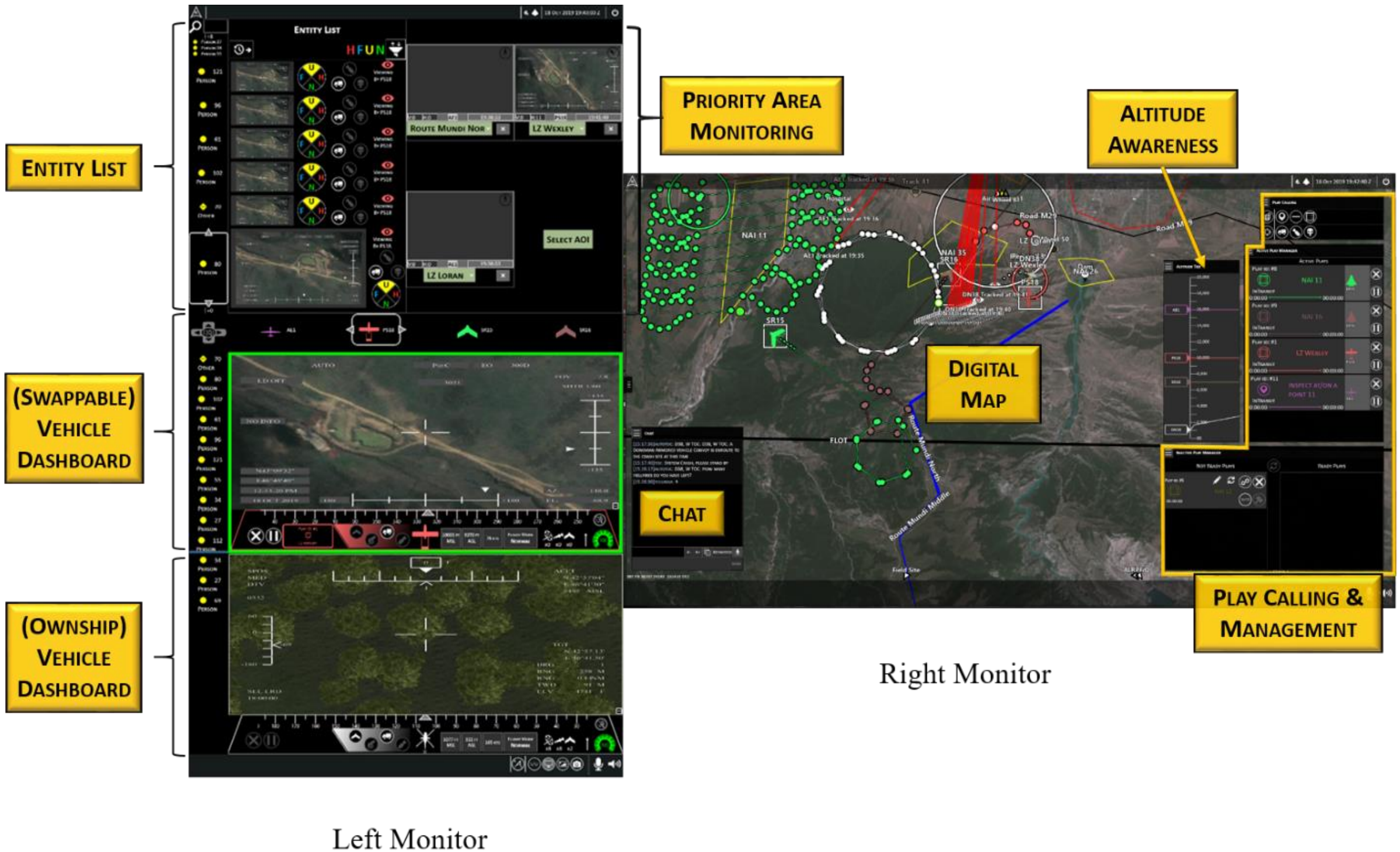


Figure 8. TECUMSA's Dual-Monitor Setup.

Participant's View "Out-the-Window" in FORCE Simulator

Figure 9 is a sample view of the FORCE simulator and the view participants had in the Round 2 Evaluation. In the image, you can see the dual monitor TECUMSA interface in the bottom left, as well as a large monitor in front of the user interface workstation that simulated what the "out-the-window" view from the Ownship cockpit would look like. Periodically participants would use this out-the-window view to help orient themselves or search for threats in the battlespace. However, the majority of participants' time was spent focusing on the TECUMSA interface.

Figure 9. Example of a Participant's View within the FORCE Simulator.

TECUMSA: User Workflow

TECUMSA's dual-monitor setup allowed participants to accomplish the tasks required in the Round 2 Evaluation using the displays seen in Figure 8. The overarching goal of participants in Round 2 was to clear threats in the battlespace to enable friendly forces to successfully move through the area. This goal required a typical workflow, involving: Finding possible threats, confirming entities were a threat, deciding whether or not to kinetically engage threats, and conducting battle damage assessments (BDA) if threat(s) were engaged. Figure 10 is a box chart that depicts the typical decision points, actions, and interface displays that could be used to accomplish the mission workflow in the Round 2 Evaluation. Additional details of TECUMSA, including the play calling process and human-autonomy task allocation, will be discussed in the following sections.

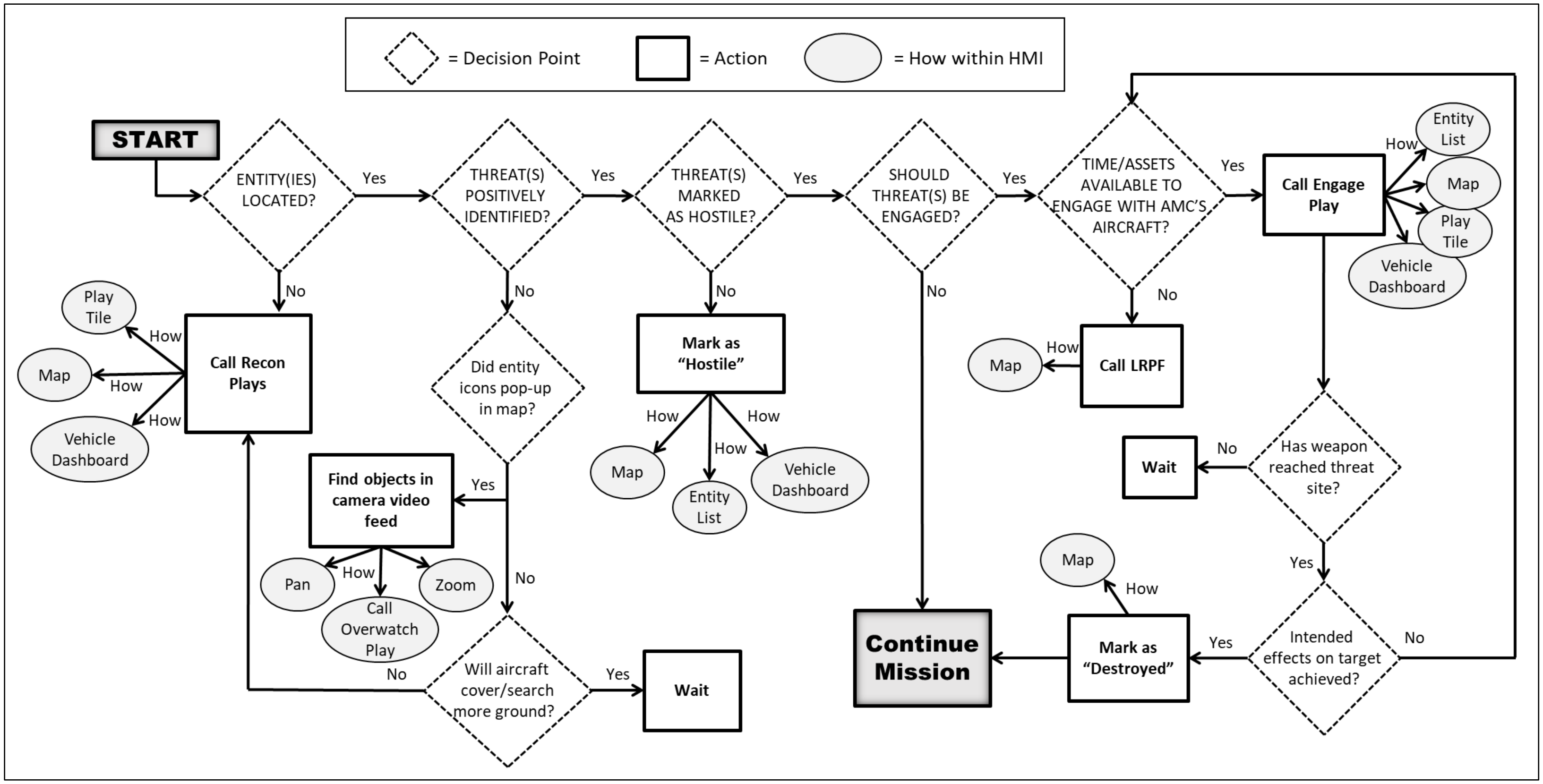


Figure 10. User Workflow in TECUMSA to Find, Identify, Track, Engage, and/or Assess Enemy Threats.

In reference to the workflow depicted in Figure 10, additional details regarding user activities are described below:

1. "Entity(ies) Located?" If not, the user should generally call a recon play and monitor the map for entity icons to pop up from ATR detections. At times the user may need to pan and zoom the camera sensor on a nearby aircraft to get the entity in the sensor's field of view.
2. "Threat(s) Positively Identified?" Using each aircraft's Vehicle Dashboard (see Figure 14), the user should view the live video feeds to look for entities that are possible threats (referencing map locations) in order to get PID. This involves the participant panning and zooming the camera sensor on a specific aircraft while cross-checking with the map to get the sensor footprint projection (field of view) to overlap with the entity icons on the map.
3. "Threat(s) Marked as Hostile?" Once a threat is found, the user could reclassify from "Unknown," to either "Hostile," "Neutral," or "Friendly." Note, entities were required to be marked as Hostile for the system to allow an engagement play to be called for that threat.
4. "Should Threat(s) be Engaged?" If yes, the user should call an Engage/Cooperative Engage Play or "Call for Fire" which uses LRPF. The user would also monitor the Vehicle Dashboard's live video feed(s) to maintain PID of the threat.
5. "Time/Assets Available to Engage with AMC's Aircraft?" The user would have to decide if the threat was dangerous enough to the mission that it was worth expending additional weapon resources and/or time.
6. "Intended Effects Achieved on Target?" The user would need to conduct BDA to decide if either the "desired effect occurred" or if "re-engage is needed." If the former, the user could mark the threat as "Destroyed," if a re-engagement is needed the user should decide when to confirm the release of additional weapons.

TECUMSA: Play Calling Process

One of the central features of TECUMSA's HMI is that it provides the operator control of the manned and unmanned aircraft through "play calling" capabilities. Much like a coach calls well-designed plays during a football game, the operator calls "plays" to support the specific tasking requirements given the unique situations imposed by the operational conditions. This play-calling approach facilitates rapid task delegation by allowing the operator to quickly execute and adapt plays in response to changing mission requirements, while the autonomous systems are responsible for optimizing aircraft allocation, route planning (e.g., avoiding NFAs, deconflicting for mountainous terrain), and continuous flight control (e.g., steering, airspeed management, waypoint following).

A series of plays that would be beneficial in an air assault mission were identified in collaboration with SMEs. Figure 11 shows the Play Calling tile, containing a list of the available plays to users in the Round 2 Evaluation. On the top row are plays focused around reconnaissance, or finding potential threats on the ground. On the bottom row are plays for keeping "eyes-on" (persistently watching) an object/threat, or kinetically engaging a threat. More specifically, the top row of plays from left-to-right (ignoring the leftmost "category" icon) are as follows:

- *Point Inspect*: Aircraft travels to a point and hovers or loiters with its sensor focused on the point of interest.
- *Route Inspect*: One or more aircraft travels along a route with their sensor focused on the route.
- *Parallel Search*: One or more aircraft searches a specified area using a bidirectional raster scan search pattern.

The bottom row in Figure 11 contains surveillance and engagement based plays, and from left-to-right (ignoring the leftmost "category" icon) are as follows:

- *Overwatch*: Aircraft continuously updates its position to ensure its sensor is fixed on the object of interest.

- *Engage*: Armed aircraft uses onboard weapon(s) to kinetically strike a target (more detail in the *TECUMSA: Human-Autonomy Teaming During Engagements* section).
- *Cooperative Engage*: Two aircraft collaborate to remotely engage a target, where one aircraft laser designates the target and the second aircraft uses onboard weapon(s) to engage it.

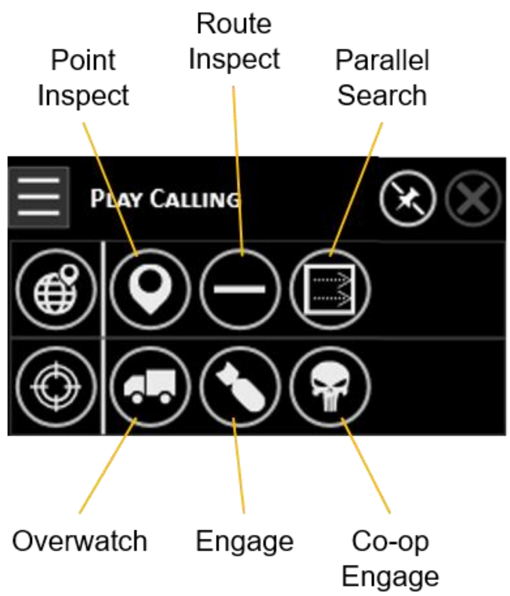


Figure 11. Play Calling Tile.

Once the user selects the play they want to execute, the Play Workbook opens (see Figure 12) giving the user the option to further customize the play through optimizations and constraints. The user could adapt a play, based on their higher-order objectives, by inputting optimizations (e.g., "speed of task completion", "combat power"), which allows the autonomy to make a more informed recommendation based on the *situational values and priorities* of the user. The Play Workbook design subtly shaped user behavior by encouraging interactions at a more efficient and effective level of control. The design of this interface component allowed the user to both clearly *observe* and quickly *direct* the solution without having to take an excessive number of actions.

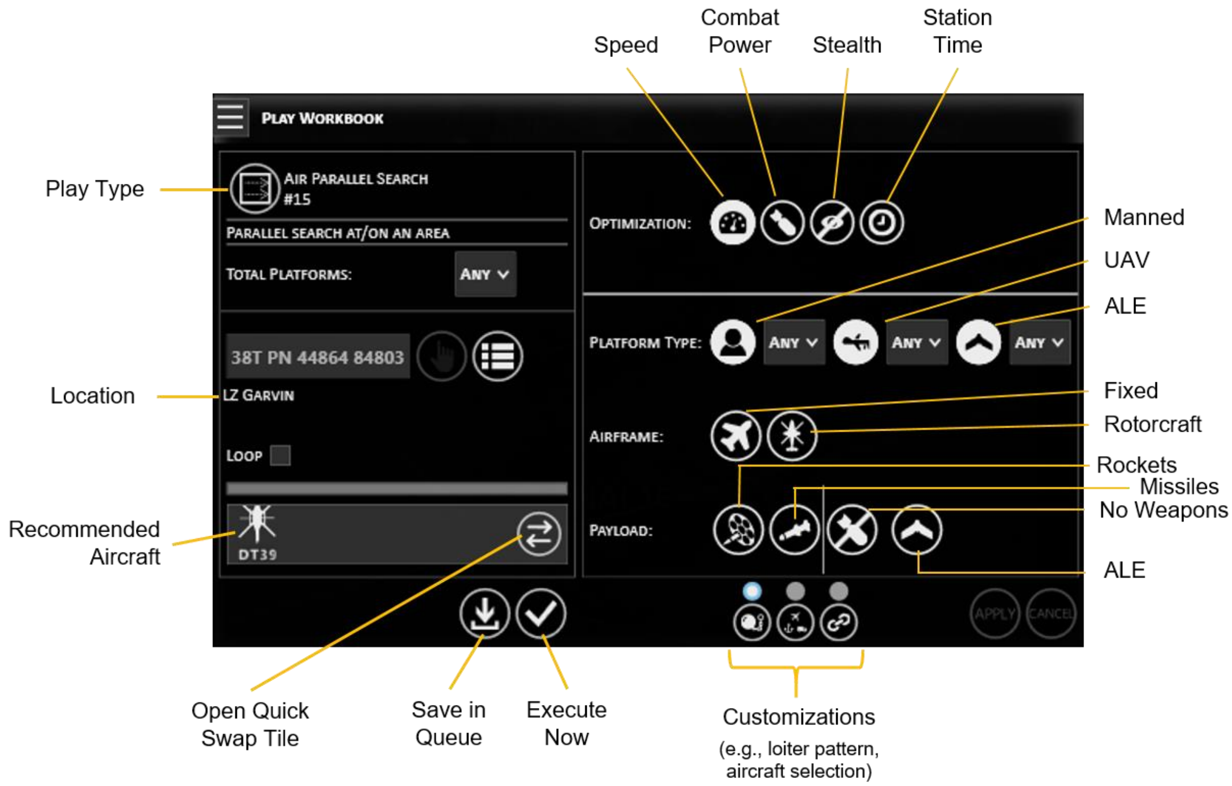


Figure 12. Play Workbook Tile.

The *left half* of the Play Workbook allows the user to communicate with the intelligent autonomous agents the type of play (e.g., Air Parallel Search, Point Inspect), the total number of platforms/aircraft needed to complete the play (e.g., Any, 2), and the location of the play (e.g., LZ Garvin). These parameters then guide the intelligent agent in its recommendation for what aircraft (and how many) to use. In the Figure 12 use case, "DT39" is recommended to complete the play. On the *right half* of the Play Workbook the user can direct the autonomy's aircraft selection based on high-level goals. These are "value" constraints that allow an algorithm to pick from alternative ways to accomplish the user's intention (i.e., determine the most satisfactory choice). Typically, this might be in terms of a variable to be minimized (time or distance) or maximized (combat power).

Additionally, the AMC can specify further criteria for accomplishing the play (e.g., designate a specific aircraft type or a specific payload).

Starting from the top-right of the Play Workbook are the “Optimizations”, including optimizing for “speed,” “combat power,” “stealth,” and “increased station time” (icons moving left to right in Figure 12). Additionally, there are hard constraints or requirements that must be satisfied by the allocator agent. The first is “Platform Type” – which specifies the minimum number of each type of aircraft required for the play, in Figure 12 “Any” number of manned aircraft, UAVs, and ALEs (Air Launched Effects) can be used. Next is “Airframe” –which specifies to the allocator agent whether “fixed-wing” and/or “rotorcraft” are required, since they each offer different capabilities in terms of maneuverability (e.g., maximum bank angle, ability to hover in place). Finally, “Payload” – specifies requirements for what is needed onboard the chosen aircraft, and moving left to right the icons indicate an aircraft with “rockets,” “missiles,” “no weapons,” and/or “ALEs” is required for the play.

At the bottom-right of the play workbook are three buttons to swap panels in the Play Workbook, to access additional customization options. This includes the ability to *specifically omit certain aircraft* to be considered for the play, setting *loiter parameters* (e.g., loiter size, pattern, standoff distance), and *chaining* or restricting the start sequence of plays. Additionally, there are two buttons at the bottom left to drop the play into a queue to access later (speeds the play calling process when several customizations are made to the play), as well as an “Execute Now” button.

Play Workbook: Quick Swap Tile

Of special note is the button to access the "Quick Swap" tile (i.e., the button located to the right of the aircraft recommendation, "DT39" in Figure 12). This button opens the tile seen in Figure 13, referred to as the Quick Swap tile. The Quick Swap tile allowed the operator to access additional logic of the allocator agent, where they could gain transparency into how particular aircraft were ranked by the intelligent allocator agent relative to one another. The Quick Swap tile was intended to help the operator understand how and why allocation plans were ranked for a given play. One main feature is the parallel coordinates plot (right half of Figure 13), which showed a line graph plot for how each plan ranked on each of the four *optimization parameters* (i.e., "speed of task completion," "combat power," "stealth," "station time").

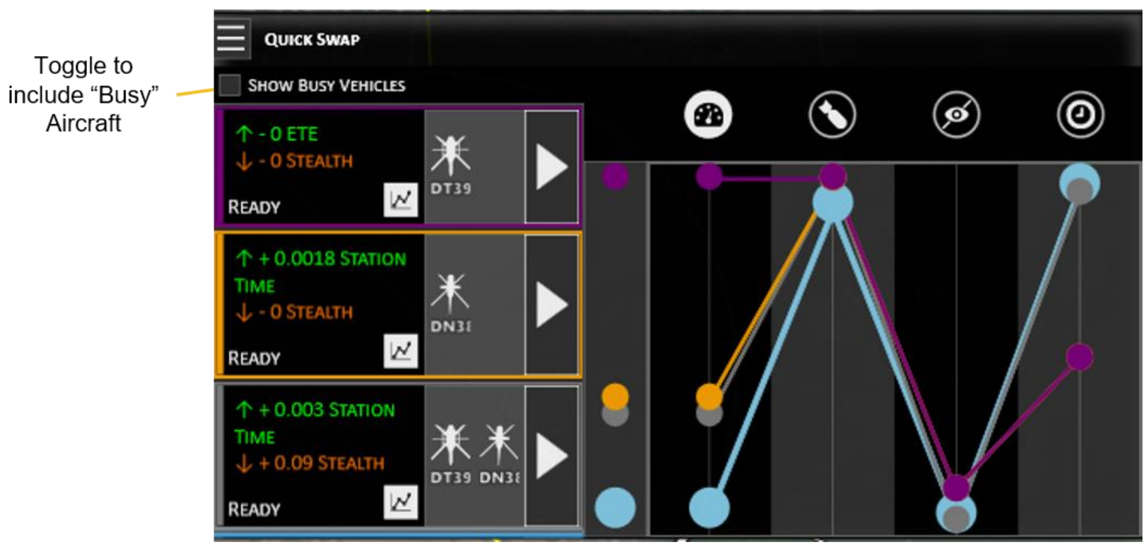


Figure 13. Quick Swap Tile Opened from the Play Workbook Tile.

It was a deliberate choice in the Round 1 version of TECUMSA to have the allocator agent consider a particular portion of the solution space when determining the "optimal" aircraft for a task. More specifically, in Round 1 the *currently available aircraft* were far preferred to aircraft that were already tasked. For our Round 2 system, however, we recognized that a user may like to view the entire solution space, regardless of current aircraft availability. As seen in Figure 13, the Quick Swap tile was made to include a selectable option to “Show Busy Vehicles.” This meant that both busy and available aircraft could be weighted equally by the allocator agent, so the optimal solution would be given regardless of aircraft availability. This feature allowed the operator to decide if it was worth pulling an aircraft off its ongoing task and re-assigning it to a new task, or if waiting for a particular aircraft to finish its ongoing task was a better option.

Another noteworthy feature in the Quick Swap tile was added after the Round 1 Evaluation, intended to support decision-making. More specifically, the feature added quick summary information to each of the plans generated by the autonomy. This information was the “biggest advantage” and “most notable loss” by choosing *this* plan (e.g., the yellow highlighted plan in Figure 13), rather than the autonomy’s most preferred plan (see the purple highlighted plan at the top of the list in Figure 13). Unfortunately, there was not enough time in the software development cycle before the Round 2 Evaluation to fully refine this feature to make the numerical values more human interpretable. However, when users are faced with a set of tradeoffs, providing concise, easily understandable summary information of the largest gain and loss can simplify decision-making and increase transparency into the autonomous agent’s reasoning.

TECUMSA: Human-Autonomy Teaming During Engagements

In the Round 2 Evaluation, the task of engaging threats involved significant human-autonomy teaming. In the synthetic task environment, after the operator finds a hostile threat, they would initiate a kinetic engagement by calling an engage/cooperative engage play. After selecting the object of the play (i.e., “Hostile Entity 444”), the autonomous system would generate a route to the threat, and once approved by the operator, the autonomy would handle the flight control tasks to navigate the UAV into weapons release range.

Once in range to fire weapons at the threat, the interface would display a “Fire Now” window in the Vehicle Dashboard of the engaging aircraft (see Figure 14). This “Fire Now” window indicates the handoff of the engagement task over to the human operator. It is the human’s responsibility to obtain positive identification of the threat and decide when to press the interface controls to actually fire the weapon, along with deciding which weapon to fire. Unless the engagement play is cancelled by the operator, the autonomy continues to keep the aircraft in weapons engagement range, as well as keep the camera sensor and laser designator locked on to the threat. This teaming between the operator and the autonomy enables the operator to offload the continuous manual control task of flying the UAVs while maintaining control over the higher cognitive function of deciding when to engage the threat.

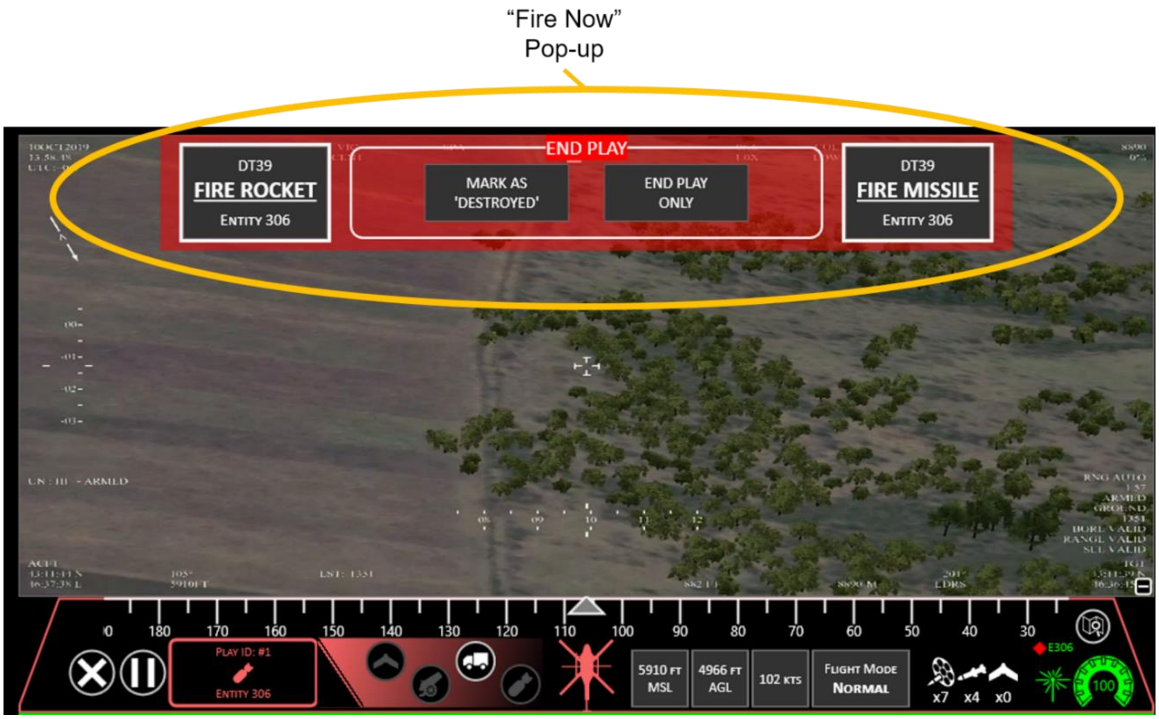


Figure 14. Vehicle Dashboard During a Threat Engagement, with Weapon Firing Options.

TECUMSA: Audio-Visual Feedback

Up to this point in the chapter, the TECUMSA interface has been discussed in terms of the visual components. However, there are also notable elements of auditory information that were part of the interface design. Radio communications back-and-forth with the TOC and Ownship pilot were a regular part of the participant's auditory information flow (via headsets). As for the TECUMSA based audio feedback, it was all fed through the participant's headset delivering a mono sound to both ears. Three of the most prevalent pieces of auditory feedback through TECUMSA that participants received in the Round 2 Evaluation were *chat messages*, *gunfire detection*, and *enemy missile launch detection*. The following is a more in-depth description of each audio type.

Chat: The Chat tile in TECUMSA offered both incoming and outgoing text-based chat messaging. An audio notification would trigger whenever an incoming chat message was received, or when a chat message was sent by the operator. The audio notification was a simple tone, using a familiar chirp sound to what is commonly heard when a new text message is received on a cellphone.

Gunfire Detection: In the Round 2 Evaluation, a simulated acoustic threat detection system would indicate to the participant when small arms weapons (hereon referred to as “gunfire”) were fired in the battlespace. When gunfire was detected, TECUMSA provided the participant with an auditory alert that was a unique tone followed immediately by an audio alert stating “Guns Guns Detected.” In addition to the audio alert, a unique icon would simultaneously appear on the map at the point of origin of the gunfire (see Figure 15 and Figure 16). As a special note, the tone used for gunfire imitated the sound of rapid gunfire, but with a synthetic sound to preserve an artificial quality since in the real-world too high of fidelity could be mistaken for actual gunshots at the Ownship.

Figure 15. Gunfire Detection Alert Icon.

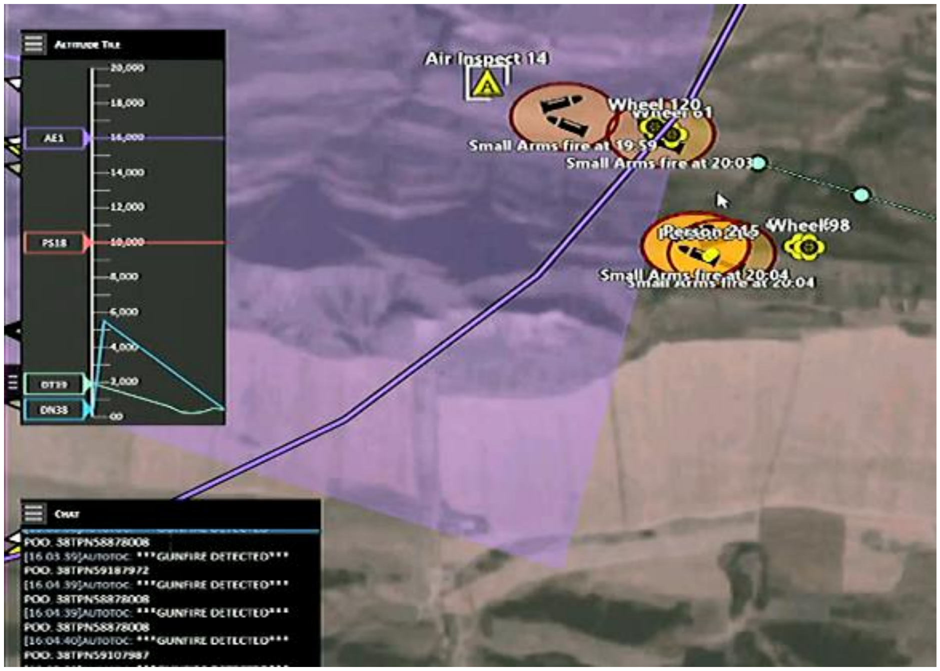


Figure 16. Gunfire Detection Alert Icons on the TECUMSA Map.

Enemy Missile Launch Detection: In the Round 2 Evaluation, simulated enemy activity included ground-launched missiles in the battlespace. In Vignette 4 an Enemy Missile Launch Detection notification indicated that enemies were firing upon one of the participant’s team of aircraft, *but not which aircraft specifically*. When an enemy missile launch was detected, TECUMSA provided the participant with an auditory alert that was a unique tone followed immediately by an audio alert stating “Missile Launch Detected.” In addition to the audio alert, a unique icon would simultaneously appear on the map at the point of origin of the enemy missile launch site (see Figure 17 and Figure 18). In preparing for the Round 2 Evaluation interface designers made the choice, based on our SMEs recommendation, to make enemy missile launches one of the most auditorily salient events in TECUMSA. This was done because of the high potential cost of losing an aircraft, both monetarily and to overall mission success. To help achieve this saliency, the auditory tone simulated the real-world sound pilots typically hear when having a missile detection event.

Figure 17. Enemy Missile Launch Detection Icon.

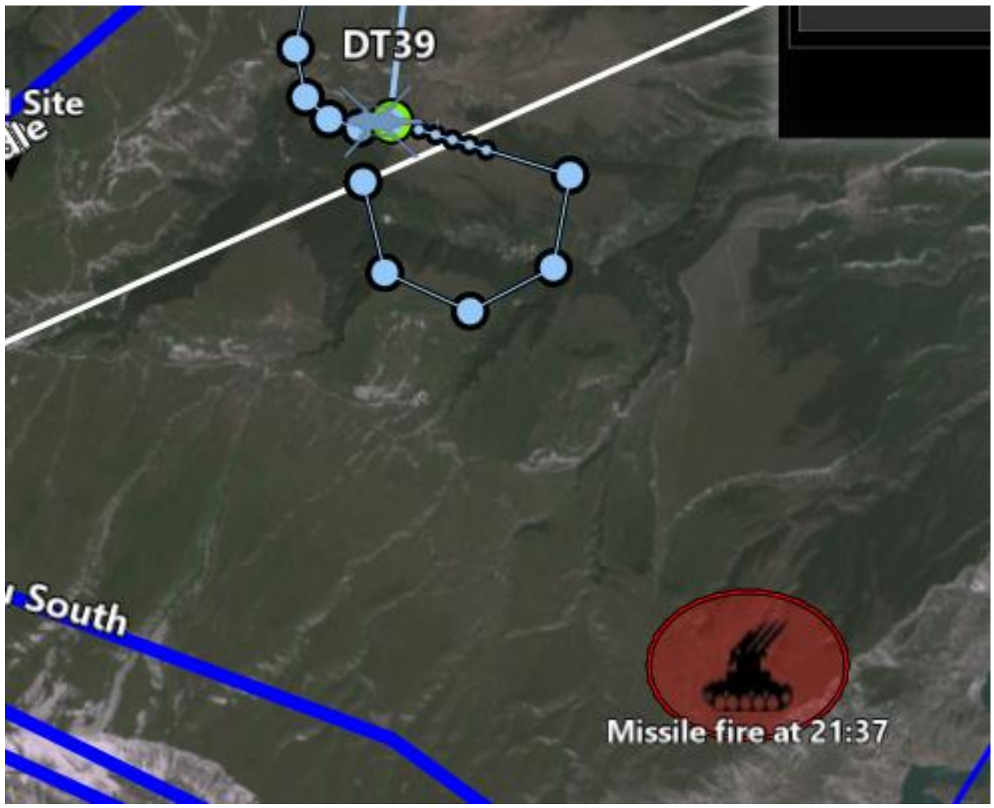


Figure 18. Enemy Missile Launch Detection Alert Icon on the TECUMSA Map.

CHAPTER 4 – Vignette 4 Timeline Narratives

The longest and most complex vignette across both Round 1 and 2 Evaluations was Vignette 4 in the Round 2 Evaluation. Not only was this vignette scripted for 60 minutes as opposed to 30 minutes, but the pop-up events had significant consequences on AMC task prioritization for the overall mission. These pop-up events created numerous key decision points for the participant during the mission execution, where the participant's course of action would have to adapt and evolve in interesting ways if they chose to address the pop-up events.

To ground the following chapters in meaningful data, the performance exploration from the Round 2 Evaluation will concentrate in detail on just two of the participant's performances during the Vignette 4 scenario. Participant 2 and Participant 7 were particularly interesting candidates to compare Vignette 4 performances. This is because both participants were part of only three of the seven participants that adequately supported the unanticipated ambush of Ground Troops by enemy forces (a pop-up event). In addition to their successful adaptation to assist the ambushed Ground Troops partway through the mission, they also had distinct approaches from one another for accomplishing the series of tasks in the Vignette 4 mission that are interesting to explore from a designer's perspective.

To better illustrate a given participant's activities and shifting priorities over time, a minute-by-minute timeline for Vignette 4 was created for "Participant 2" and also "Participant 7." Moving left to right in Figure 19, this timeline reveals a minute-by-minute breakdown of the given participant's unique goals, aircraft tasking choices, threat strikes/engagements, as well as enemy activity in the battlespace that frequently guided their chosen course of action. The next section will provide a more in-depth description of the different elements displayed in the timeline illustrations. Following that, a walkthrough of Participant 2 and Participant 7's unique courses of action in Vignette 4 will be discussed and then compared.

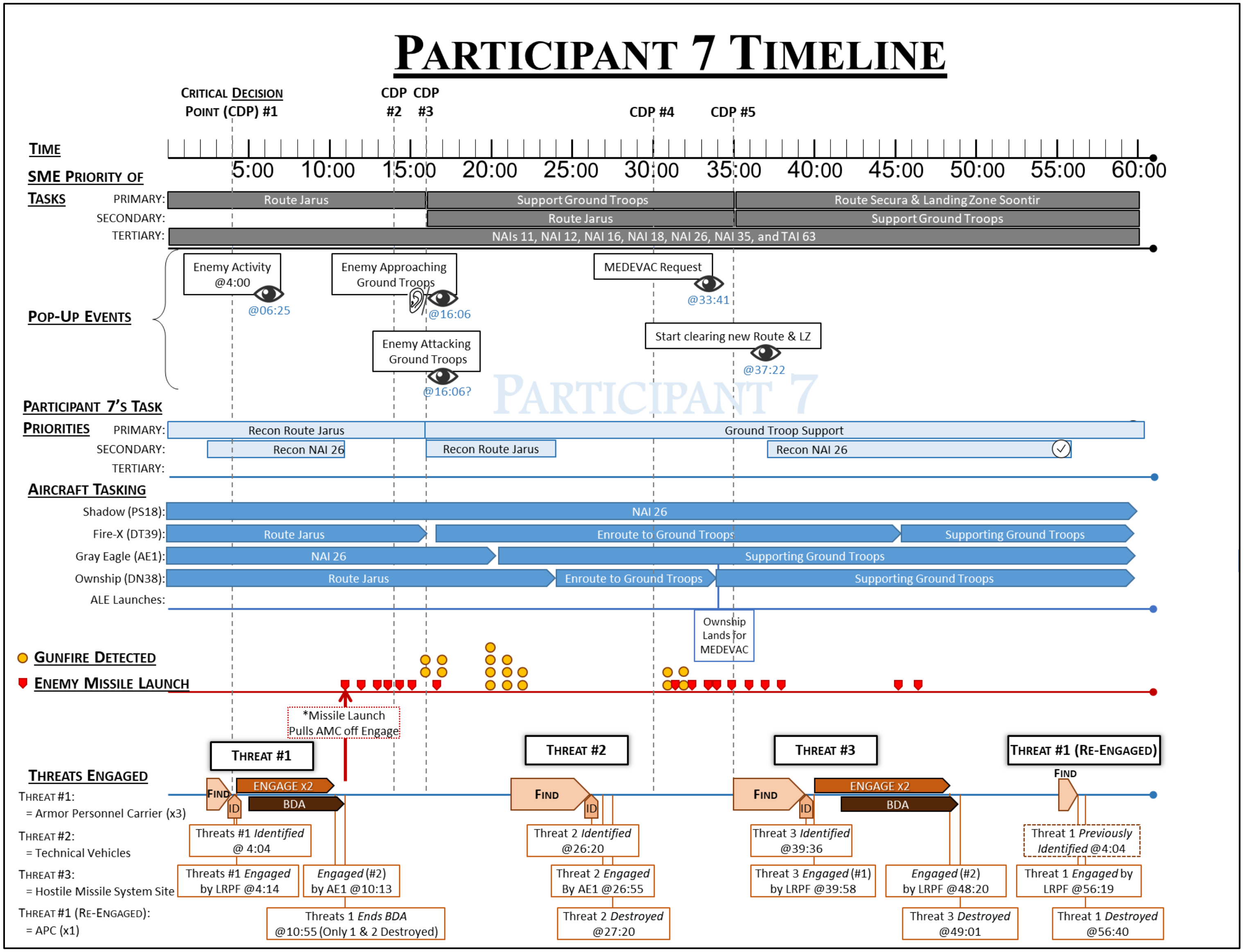


Figure 19. Participant 7's Timeline of Tasks and Priorities for Vignette 4, Round 2 Evaluation.

Timeline Figure Overview

For two participants in the Round 2 Evaluation (Vignette 4) an illustration of the minute-by-minute timeline from their mission execution was created (see Figure 19 and Figure 20). Included in each participant's timeline are the following pieces of information:

- *Time*: Minute-by-minute timeline, which starts the 60-minute vignette on the left and ends on the right.
- *SME Priority of Tasks*: Corresponds to the SME verified priorities that participants should have had for the tasks required in Vignette 4. Based on the information provided to participants, at any given time there could be up to three levels of task priority: *Primary Task* (highest importance), *Secondary Task*, and *Tertiary Task* (lowest importance).
- *Pop-up Events*: Pre-scripted events or information transmissions, which were not known ahead of time by the participant, and intended to add complexity to the vignette and potentially change their originally planned course of action as an AMC. In Vignette 4 for example, roughly 16 minutes into the mission a pre-scripted ambush of Ground Troops occurs by ground threats in the area.
- *Critical Decision Points*: These scripted pop-up events correspond to "Critical Decision Points," where the participant could/should have chosen to adjust their course of action based on the new information.
- *Participant's Priority for Tasks*: The breakdown of a given participant's priority of tasks in Vignette 4. This section of the timeline illustration allows easy comparison to be made between the Priority of Tasks determined by SMEs ahead of time, and the Participant's Priority of Tasks throughout the mission for each of their aircraft.
- *Aircraft Tasking*: The participant's minute-by-minute tasking of each available aircraft. The participant had four aircraft in-flight (i.e., Shadow, Fire-X, Ownship, Gray Eagle), and four small UAVs hereon referred to as "ALEs." The ALEs were unique in that they were carried *onboard* two of the aircraft (see

Table 3 for each aircraft's payload and number of ALEs onboard). Because the ALEs were carried onboard the Ownship and Gray Eagle, the participant could only use their camera sensors once they *launched* the aircraft (launch command issued through the Vehicle Dashboard). ALEs are included in the list of Aircraft Tasking in Figure 19 and Figure 20 if that participant chose to launch them. Table 3 provides details for the payload of each aircraft available to the AMC/participant.

- *Gunfire Detection*: When gunfire was detected, TECUMSA provided the participant with an auditory alert, as well as populated a unique icon in the map at the point of origin of the gunfire. More details on audio-visual feedback of gunfire detection alerts are provided in the *TECUMSA: Audio-Visual Feedback* section.
- *Enemy Missile Launch Detection*: When an enemy missile launch was detected, TECUMSA provided the participant with an auditory alert, as well as a unique icon in the map at the point of origin of the enemy missile launch site.
- *Threats Engaged*: Indicates when a participant kinetically engaged enemy threats. The typical stages of an engagement are as follows: 1) *Find* possible threats, 2) *Identify* to confirm PID of a threat, 3) *Engage* threat by firing missile/rocket off one of the aircraft, or by calling for a missile to be fired from a remote ground site in the battlespace (referred to as an LRPF), and 4) *BDA* to confirm the neutralization/destruction of the threat, and *reengage threat* if needed.

Table 3. Aircraft Payload in Vignette 4, Round 2 Evaluation.

ROUND 2, VIGNETTE 4

	Shadow (PS18)	Fire-X (DT39)	Gray Eagle (AE1)	Ownship (DN38)	ALEs (SR15 – SR19)
Camera Sensor	Equipped	Equipped	Equipped	Equipped	Equipped
Max Airspeed (knots)	70 kts	115 kts	134 kts	250 kts	80 kts
Radar Warning Receiver (aids threat detection)	N/A	Equipped	Equipped	Equipped	N/A
Missiles	–	4	2	4	–
Rockets	–	7	–	7	–
ALEs Onboard	–	–	2	2	–

*Note: All values are notional and for simulation purposes only, they are not reflective of any actual fielded weapon systems.

Participant 7's Timeline

Due to the different density of events in each participant's timeline, Participant 7 will be discussed *before* Participant 2 so additional intricacies can be built upon in the latter timeline narrative. First in this section is a detailed narration of Participant 7's Vignette 4 performance, further describing the events depicted in the timeline seen in Figure 19. Following the detailed narration for Participant 7 is a summary of their observed performance. As a special note, *the response timings discussed here are notional and for simulation purposes only, they are not reflective of any actual fielded weapon systems.*

Participant 7's Detailed Timeline Narration

Participant 7 begins their course of action by having their "Fire-X" UAV and Ownship reconning Route Jarus, and their "Shadow" and "Gray Eagle" UAVs reconning NAI 26; these aircraft remain on these tasks for the first quarter of the 60-minute vignette. At 04:00 minutes into the vignette, Participant 7 receives a chat message with information from the first pop-up event reporting heavy enemy activity near the Ground Troops to the East of the dam. At 04:04 the participant positively identifies a group of hostile enemy armored personnel carriers (APCs) in NAI 26, which they choose to kinetically engage twice over the next seven minutes. Although the participant's attention was primarily focused on engaging the hostile APCs, mid-way through this engagement the participant briefly turns their attention over to read their chat message (at 06:25), 2 minutes and 25 seconds after receiving it.

At 14:00 minutes into the vignette, the Ground Troops report via chat that an enemy patrol is approaching their position (the second pop-up event), which the participant reads 2 minutes and 6 seconds later. The previous two chat messages (at 04:00 and 14:00 minutes) were foreshadowing of the next chat message at 16:00, which is a critical turning point in the scripted mission priorities. At 16:00, the Ground Troops report via chat that they are now under attack (the official start of the third pop-up event which is the most mission impactful), and request immediate armed assistance from the participant. Simultaneous to this incoming chat message, two audio/visual alerts of gunfire detection occur, along with a large orange icon in the map indicating the precise location of a gunfight that the Ground Troops are now in with an enemy patrol.

Up until the time of the gunfire detection at 16:00, there have been 6 missiles launched at one or more of the participant's team of aircraft. Note the participant is not told which aircraft is getting shot at, only the grid location the enemy missile launched from. In the TECUMSA system, enemy missile launches were accompanied by both a salient auditory warning as well as a large icon in the map indicating the location of the enemy missile launch. Despite the saliency and time pressure of these missile launches, the audio/visual alert of the gunfire was prominent enough to grab Participant 7's attention. Shortly after noticing the gunfire, Participant 7 quickly tasks Fire-X to provide support to the Ground Troops.

At 20:41 the participant realizes how long it would take for Fire-X to reach the Ground Troops, and decides to also task the Gray Eagle to provide armed support to the Ground Troops under attack. Because the Gray Eagle was already in range of the Ground Troops, the participant was able to immediately start searching for entities in the location of the gunfire in hopes of finding the threats attacking the Ground Troops. After searching the map and video feeds for entities near the gunfire for 3 minutes and 27 seconds, the participant finds a few wheeled vehicles in the location of the gunfire but has not yet been able to positively identify them as friendly or enemy.

Around 24:00 the participant notices that Fire-X is still several minutes away from the Ground Troops. Since Gray Eagle has only 1 missile remaining in its weapons payload, the participant shifts their attention to getting additional armed support over to assist the Ground Troops. At 24:08 the participant pulls the Ownship off the task of reconning Route Jarus and instead directs it to head over to support the Ground Troops, which becomes their third of four available aircraft (excluding the ALEs, which are not launched) tasked to assist the Ground Troops up to this point in the vignette. After tasking the Ownship, the participant returns to looking for the enemies engaging the Ground Troops through Gray Eagle's camera sensor.

At 26:20 the participant is able to positively identify the three hostile enemy technical vehicles attacking the Ground Troops, and then reclassifies the threats as "hostile" in the TECUMSA system. At 26:30 the participant calls an Engage Play using the Gray Eagle, and at 26:55 the participant launches missiles off of Gray Eagle at the hostile threats. At 27:20 the participant confirms the hostile technical vehicles are destroyed and has therefore successfully supported the unanticipated pop-up event of ambushed Ground Troops.

After neutralizing these threats, the participant returns to trying to precisely locate the Ground Troops location and communicates with the TOC (played by an experimenter) to get information on the status of the Ground Troops. The participant specifically asks the TOC over radio (via headsets) communication, "Are they [Ground Troops] alive?" Behind the scenes the TOC stalls to answer the participant's question, because in 55 seconds the fourth critical chat message is scheduled to be sent to the participant (at 30:00), requesting a medical evacuation (MEDEVAC) for the Ground Troops. In the meantime, as the participant waits for status information on the Ground Troops, they work to find threats from the surrounding areas of the Ground Troops' approximate location.

At 30:00 the fourth pop-up event initiates, where a chat message is sent to the participant requesting MEDEVAC support for the Ground Troops. At this point, the Ownship is still enroute to the Ground Troops. At 33:41 the participant checks their chat window and sees the Ground Troops' request for a MEDEVAC. By the time the Ownship reaches the Ground Troops location, at 34:01, four missiles have been launched at one or more of the participant's aircraft. It is at this arrival time at 34:01 that the experimenters notice the Ownship aircraft hovering just above the ground and barely moving forward. This was because, upon arrival to the Ground Troops, Participant 7 had the pilot position the Ownship to land and actually carry out the MEDEVAC. It is worth noting that no other participant did this, and behind the scenes SMEs were extremely impressed by the participant's dedication to help the injured Ground Troops.

After the participant offers to provide the Ground Troop's MEDEVAC, 59 seconds go by. The fifth and final pop-up event, which is a scheduled chat message, is then sent at 35:00 stating that a MEDEVAC aircraft team is 25 minutes from the current Ground Troop's location and will fly ingress via Route Secura and land in LZ Soontir. Since Participant 7 has their Ownship positioned to actually carry out the MEDEVAC, the TOC adapts the script by further clarifying to the participant over radio communications to "Charlie Mike" (or Continue Mission) because the Ground Troops need specialized medical equipment.

At 35:00 the participant turns their efforts towards clearing the surrounding area of threats. After 4 minutes and 36 seconds of searching, the participant positively identifies an anti-aircraft missile system (hereon referred to as a "hostile missile system"), which has been launching several missiles at the participant's aircraft throughout the mission. At 39:58 the participant sends the first LRPF at this threat (i.e., calling for a missile launch from a remote ground location), and after not observing the desired kinetic effects, the participant decides at 48:20 to send a second and final LRPF at the hostile missile system. Shortly after the second LRPF, the participant confirms the hostile missile system is destroyed through a nearby aircraft video feed and marks it as destroyed in the TECUMSA map.

At 49:00 into the vignette, 14 minutes after being notified of the new route and landing zone needing to be reconned (i.e., Route Secura and LZ Soontir) for the Ground Troop MEDEVAC, the participant still has three of the four aircraft loitering nearby to the Ground Troops. In other words, the participant has not tasked any of the aircraft to specifically search along Route Secura or LZ Soontir. At 56:16 the participant finds one of the APCs in NAI 26 that had not previously been impacted by the first engagement. At 56:19 the participant sends an LRPF request, and at 56:40 confirms the APC threat has been destroyed.

At 60:00, the end of the vignette, the TOC asks the participant whether or not they advise the MEDEVAC operation to continue. Since the participant had neutralized the known hostile missile system and did not detect or know of any other threats that would impact the incoming MEDEVAC team, they gave the TOC an affirmative "Go" call to continue with the MEDEVAC. After consulting with SMEs after the completion of this run, both SMEs observing the vignette agreed that this was a good call made by the participant.

Participant 7 Summary

Participant 7 was one of the top performers in their armed support of the ambushed Ground Troops, despite this being an unanticipated pop-up event. Participant 7 demonstrated a tasking strategy that utilized redundancies in aircraft tasking, with three of the four main aircraft tasked to support the Ground Troops for roughly 75 percent of the mission. On the plus side, this tasking strategy ensured the Ground Troops were absolutely supported by armed aircraft. The negative of this approach is that it resulted in only 3 of the 11 reconnaissance tasks being addressed in Vignette 4.

One of the most noteworthy choices of the participant was their decision to *not* launch any of their four available ALEs, which would have provided four additional aircraft with video feed coverages of the battlespace. It is worth noting that Participant 7 *had* launched ALEs in prior vignettes that day. Despite never launching any of the available ALE aircraft, Participant 7 was still able to eliminate four of the seven high-priority threats in Vignette 4 (Round 2). At the end of each vignette the participants are asked by the TOC whether the following mission is a “Go,” or “No-Go.” This question boils down to the participant essentially being asked “Did you sufficiently search and clear the necessary threats for the incoming aircraft to fly through safely?” Because Participant 7 neutralized the known threats in the battlespace, they confirmed with the TOC that the pop-up MEDEVAC could continue (i.e., “MEDEVAC is a go”), which the two SMEs observing the performance agreed was a good call.

Participant 2’s Timeline

The next section is a detailed narration of Participant 2’s Vignette 4 performance, further describing the events depicted in the timeline seen in Figure 20. Following the detailed narration for Participant 2 is a summary of their observed performance. Note, *the response timings discussed here are notional and for simulation purposes only, they are not reflective of any actual fielded weapon systems.*

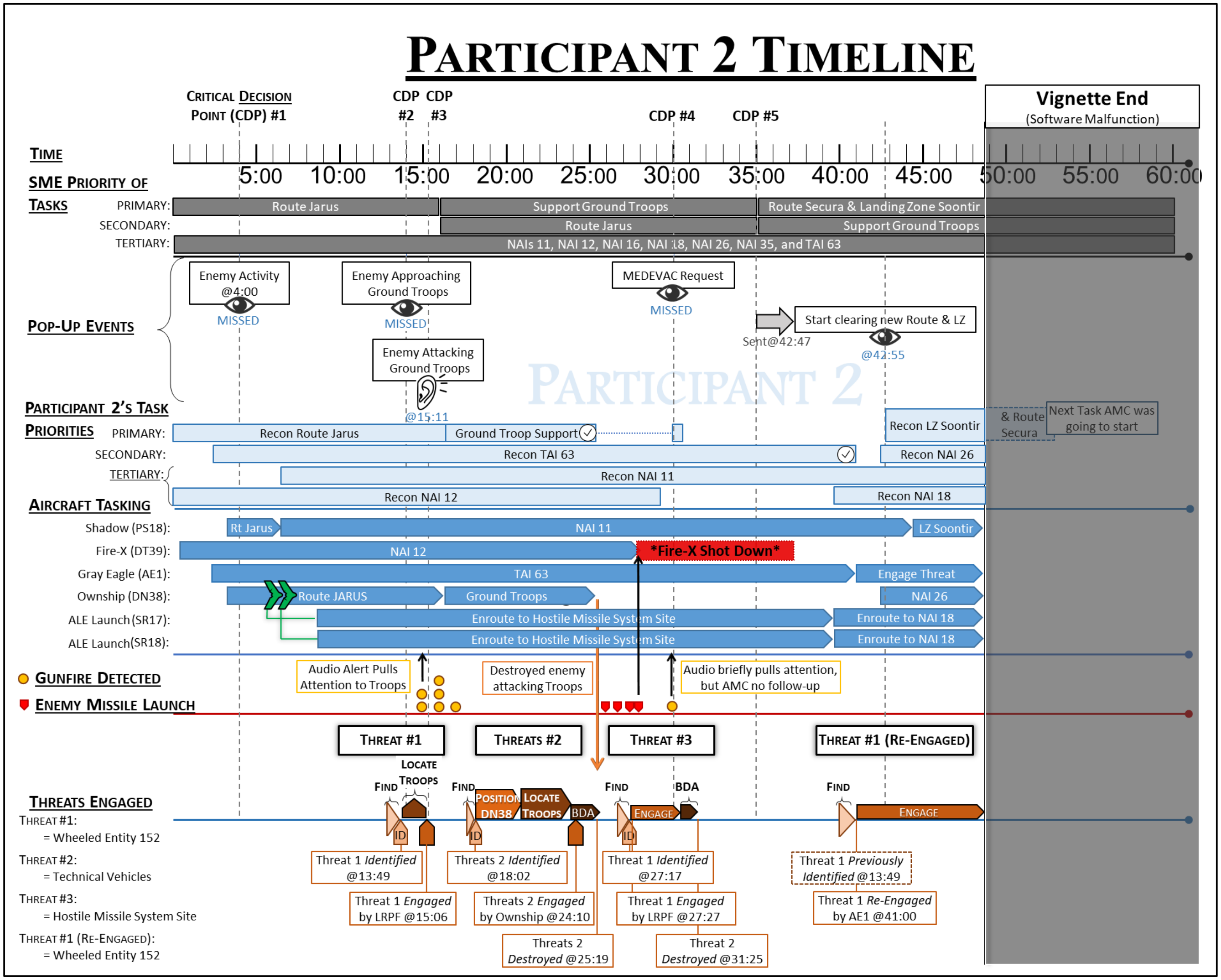


Figure 20. Participant 2's Timeline of Tasks and Priorities for Vignette 4, Round 2 Evaluation.

Participant 2's Detailed Timeline Narration

In the first five minutes of Vignette 4, Participant 2 has their "Fire-X" UAV tasked to recon NAI 12, their "Gray Eagle" UAV reconning TAI 63, and their "Shadow" UAV and Ownship tasked to do a joint recon of Route Jarus. At 04:00 minutes into the vignette, Participant 2 receives a chat message pertaining to the first pop-up event reporting heavy enemy activity near the Ground Troops to the East of the dam, but the participant never read this chat message. At 06:29, the participant pulls the Shadow off of the joint recon of Route Jarus after noticing that it will take a significant amount of time for the Shadow just to travel across the battlespace and start reconning Route Jarus. The participant then instead tasks the Shadow to recon NAI 11, which was a much closer area to the starting location of the Shadow.

At 06:42 the participant launches the first ALE off of their Ownship, and then at 06:51 launches the second ALE off of the Ownship. With the two available ALEs on the Ownship launched, the participant then tasks both aircraft to inspect the location of a templated hostile missile system. Once the two ALEs are tasked, the participant turns their focus to finding hostile threats in the battlespace.

At 13:49 the participant identifies a wheeled vehicle near the vicinity of the templated hostile missile system. At 14:00 minutes into the vignette, the Ground Troops report via chat that an enemy patrol is approaching their position (the second pop-up event), which the participant never reads. After confirming no friendlies are in the nearby location to the templated hostile missile system, the participant calls an LRPF on a wheeled threat at that location at 15:06. At the exact same time (15:06), a gunfire detection audio/visual alert is generated. Five seconds later at 15:11 the third critical chat message is sent, reporting that the Ground Troops are now under attack and request immediate armed assistance from the participant (the official start of the third pop-up event, which is the most mission impactful).

Although the participant does not immediately read the third chat message reporting the Ground Troops under attack, the audio/visual alert of gunfire at 15:06 was salient enough to draw their attention to the Ground Troops location on the map. Up until this point, the participant has not read any of the critical chat messages tied to the scripted pop-up events in Vignette 4. However, Participant 2 correctly decides that the gunfire is potentially coming from and/or directed towards the Ground Troops that were known to be roughly in the area of the gunfire, based on their pre-mission briefing.

At 16:13 the participant has the Ownship head towards the gunfire, but at a safe distance to the South. After redirecting the Ownship and pulling it off of the Route Jarus reconnaissance, the participant checks their chat tile for additional information in regards to the gunfire detection. At 17:10 the participant reads that the Ground Troops are requesting immediate armed support from the participant. At 17:24 the participant sees multiple wheeled vehicle icons pop up on the map at the location of the gunfire. Participant 2 then spends the next 38 seconds confirming that the newly detected wheeled entities are indeed hostile and not the Ground Troops fighting back.

From 18:02 to 20:43 the participant is directing the Ownship pilot to get in position to engage the hostile threats attacking the Ground Troops. From 20:43 to 23:04 the participant is confirming the location of the Ground Troops to make sure they are not in close proximity to the intended kinetic strike location. At 24:10 the participant calls an Engage Play using the Ownship, and at 24:30 the participant launches missiles off of their Ownship at the hostile threats. At 25:19 the participant confirms the hostile technical vehicles are destroyed and has therefore successfully supported the unanticipated pop-up event of ambushed Ground Troops.

Following this engagement, at 25:53 a missile is launched at one of the participant's aircraft. At 26:25 the participant directs their attention to the location of the missile launch site. At 27:17 the participant identifies the missile launch site as hostile and at 27:27 sends a request for LRPF at the location of the enemy missile launch. From 25:53 to 27:49, four missiles in total are launched at the participant's aircraft. At 29:12 the participant learns that the Fire-X UAV aircraft has been shot down by an enemy missile (i.e., one of the enemy missile launches hit its target). At 30:00 the fourth pop-up event initiates, where a chat message is sent to the participant requesting MEDEVAC support for the Ground Troops.

With Fire-X no longer available for tasking, the participant returns to their task of finding threats in the battlespace. From 31:58 to 37:50 the participant focuses on finding threats in TAI 63. At 37:50 the participant reorients their attention back to templated hostile missile system. At 38:51 they focus specifically on the icon in the map for Wheeled Entity 152, which is in proximity of the templated hostile missile system. However, in actuality, the Wheeled Entity 152 has moved from that location, and the color faded icon in the map only indicates the last known location. The participant misinterprets what the icon in the map indicates and decides to pursue an engagement of the Wheeled Vehicle 152 due to its proximity to the templated threat (i.e., the hostile missile system). At 41:00 the participant calls an Engage Play on Wheeled Entity 152 using the Gray Eagle.

At 42:47 the fifth and final pop-up event initiates, where a chat message is sent to the participant stating that the MEDEVAC team is 25 minutes from the current Ground Troop's location and will ingress via Route Secura and land in LZ Soontir. Eight seconds after receiving the chat message the participant reads it, and through the think-out-loud procedure states that they need to locate LZ Soontir. However, before shifting their attention to LZ Soontir, the participant checks back in with Gray Eagle's progress on the engagement of Wheeled Entity 152. The participant monitors and controls Gray Eagle's camera sensor until 44:30, at which time they then decide to pivot the Shadow's efforts from NAI 11 to instead begin addressing the need to recon LZ Soontir.

At 44:35 the user interface begins to lag in response to the participant's inputs, due to unforeseen artifacts of the synthetic world environment. Although verbally stating their task priorities, the participant struggles to work around the delayed system feedback. The system lag continues to be a challenge for the participant, and at 48:27 the experimenters choose to end the vignette. Thankfully this run through Vignette 4 was sufficiently long enough to capture the participant's chosen course of action and adaptation strategies to all of the scripted pop-up events.

At 48:27, the end of the vignette, the TOC asks the participant whether or not they advise the MEDEVAC operation to continue. Since the participant had neutralized the known threats in the battlespace, they confirmed with the TOC that the incoming MEDEVAC team could continue their mission, but advised of a known radar detection system to the North of the Ground Troops. After consulting with SMEs after the completion of this run, both SMEs observing the vignette agreed that this was a good call made by the participant.

Participant 2 Summary

Participant 2 was another one of the top performers in their armed support of the ambushed Ground Troops, despite this being an unanticipated pop-up event. Participant 2 demonstrated a very distributed tasking strategy of their aircraft, which allowed them to address 9 of the 11 reconnaissance tasks necessary in Vignette 4. The participant demonstrated a strong ability to simultaneously monitor and manage six aircraft in a dynamic, contested battlespace. This participant left no UAV untasked for more than a few minutes, really highlighting their ability to intelligently allocate their attention across the battlespace and not succumbing to tunnel vision on any one particular task.

Although Participant 2 wisely chose to launch two of their four available ALEs, the early launch times meant that it was up to the ALEs to traverse the battlespace. Because the ALEs were relatively slow-moving in the simulation, it meant the ALEs spent the majority of their time traversing to their objectives and only moderately helped accomplish reconnaissance tasks. Participant 2 was able to eliminate four of the seven high-priority threats in Vignette 4 (Round 2). Because they neutralized the known threats in the battlespace, they confirmed with the TOC that the follow-on MEDEVAC mission could continue, but advised of a known radar detection system to the North of the Ground Troops. The two SMEs observing Participant 2's performance agreed this was a good call.

Participant Comparison

Based on the information they provided in their demographics questionnaires at the beginning of the evaluation, there is a noteworthy comparison between Participant 7 and Participant 2's amount of experience. Participant 7 was 33 years of age, had 295 total hours of flight experience, 0 hours of experience as AMC, and 0 hours of combat experience. As for Participant 2, they were 39 years of age, had 3500 total hours of flight experience, 1800 hours of experience as AMC, and 2000 hours of combat experience.

There are potential variances in participant strategies due to their differences in tactical experience. Some of the key differences between these participants were observed in their aircraft tasking strategies (e.g., distribution of aircraft tasking, use of ALEs), their responses to offensive behavior of threats (e.g., missile launches), and their maintenance of situation awareness (e.g., frequent scanning of battlespace). Additionally, their prioritization of tasks was most notable. Participant 7 did not prioritize reconning the NAIs/TAIs, and instead focused on the primary mission objectives (e.g., Route Jarus and the Ground Troops). Participant 2 on the other hand, sought to achieve more widespread reconnaissance coverage of the battlespace. However, both participants demonstrated a valuable ability to adapt their course of action as new information was received, and ultimately they were able to complete the most important pop-up task of supporting the ambushed Ground Troops.

Summary

A significant part of being able to take away meaningful lessons from an evaluation is exploring the data. In lieu of a more classical controlled experimental design, the current research used a naturalistic, descriptive analysis. The first step of this analysis was to therefore map out the sequence of actions participants made within a high-fidelity synthetic task environment. In this section, two participants' performances were narrated based on the timeline of events from their Vignette 4 (Round 2 Evaluation) run. Using these same event narratives, the following chapters contain a more extensive exploration of Participant 7 and Participant 2's performance relative to *observability* and *directability* in the system design

CHAPTER 5 - Observability

Introduction

The goal for the HMI of TECUMSA was to provide the operator/AMC with the necessary situation awareness and interface usability to enable the execution and management of air assault mission tasks, including command and control of a team of UAVs and manned aircraft. The focus of this chapter is therefore on the elements of the TECUMSA interface that afforded the user the ability to observe, predict, and understand the state of the world relative to their goals and intentions, also referred to as *observability* (McDermott et al., 2018). This includes designing observability into the possibilities of action, consequences of action, status feedback, as well as transparency into the behaviors of the autonomous capabilities (Sarter, Woods, & Billings, 1997; Miller, 2014). The following writing provides an overview of observability in interface design, followed by a discussion of observability relative to the participant narratives from the previous chapter.

Designing Observability

> "*Humans do not have stable input-output characteristics that can be studied in isolation...When the system is put to work, the human elements change their characteristics; they adapt to the functional characteristics of the working system, and they modify system characteristics to serve their particular needs and preferences*" - Rasmussen, Pejtersen, & Goodstein (1994, p.6)

This opening quote from Rasmussen et al. (1994) highlights the symbiotic relationship between perception and action, where perception is an active, rather than passive process. That is, what an observer can 'see' often depends on them actively searching or looking for information. In the case of the TECUMSA interface, one aspect of this information gathering occurs when swapping between displays for different UAVs in the Vehicle Dashboard tile. Additionally, attention must be directed across the selected displays to gather a variety of pertinent mission information.

For an operator working in a complex socio-technical system to make informed decisions and set appropriate goals, they must have continuous information regarding pertinent system status or state variables. The concept of "observability" is that there are multiple layers of higher-order properties such as constraints (performance limits) that allow the operator to anticipate the possibilities for actions as well as the potential consequences of those actions (e.g., the imminence of a collision relative to braking capabilities). *Interfaces must therefore be designed to indicate the meaningful critical action boundaries relative to the operator's goal* (Hutchins, Hollan, and Norman, 1985; Norman, 1986).

Critical action boundaries include representations for *transitional states*, where shifting capabilities or constraints create a different set of action possibilities (e.g., the loss of an armed unmanned aircraft on the team). Additionally, *boundary conditions* or violations to "normal functioning" within the system should also be made observable to operators. For example, an aircraft flight path that routes into range of an enemy threat, or a sensor reading that triggers an aircraft stall warning.

The idea of observability is also pertinent for team collaborations, including teams composed of both humans and autonomous agents. Successful collaboration and coordination require a shared understanding of each team member's goals, intentions, responsibilities, priorities, progress, and difficulties (i.e., competency boundaries) (McDermott et al., 2018; Hutchins, Cummings, Draper, & Hughes, 2015). Observability in the context of human-autonomy teaming should therefore provide transparency into not only the *current* goals and progress of team members, but insights into likely *future* goals and challenges for the team as well (Klein, Woods, Bradshaw, Hoffman, & Feltovich, 2004).

However, it is not merely enough to know what needs to be done, as a user must also be given the tools to take action to control the desired state of the system (the principle of "directability", which will be discussed further the next chapter). Observability, therefore, includes the need to provide transparency into the *possibilities* and *constraints* on action, and just as importantly the *consequences* of action. In other words, not only does a user need to know that the knob is adjustable, but what happens when the knob is adjusted one way or the other.

Observability can be provided through the use of culturally appropriate metaphors, such as the trash can icon which signals the ability to remove/delete an item. Observability can also be supported in interface design through interaction ability with an item, such as the tooltip information that is presented when a cursor hovers over an item for a period of time. Ultimately, observability can take many forms (e.g., images, labels, colors), but the *forms* chosen must be guided by the *functions that necessitate observability in the first place.* After all, the need for a trash can icon is necessitated first by the need to delete something. This highlights the critical step in cognitive systems engineering, to determine what information must be exchanged and interpreted by the user for them to make decisions to accomplish their goals. Because observability of action potentials within the system ultimately shapes the user's capabilities and constraints, information prioritization must be grounded in the operational goals and tasking requirements of the work domain. Simply put, understanding the operational requirements (function) will guide the ideal visualizations (form).

Observability in Design for Participant 7

The participant narratives discussed in Chapter 4 were meant to provide an overview of the activities of the participant during Vignette 4. However, there is a deeper level of analysis worth exploring. As discussed previously, it is the coupling of perception and action that is critical to understanding whether information is well specified and whether the participant is well-tuned to the necessary affordances of the work. In subsequent sections, participant tuning to these affordances will be discussed.

In this chapter specifically, the focus will be on tuning (mis)alignments that were related to *observability*. For example, were audio alerts salient enough to shift the participant's focus to a higher urgency task? Did the representation of information enable the participant to see the solution space in terms of the *possibilities for action* and *consequences of (in)action* **(**e.g., not launching any ALEs)? Was the participant able to perceive or anticipate constraints, such as performance limits (e.g., max airspeed), or *transitional states* where the opportunities or consequences/risks of actions shifted (e.g., a shift in priority once Ground Troops were under attack, dwindling weapon payload on GE1).

Three major elements of observability will be discussed in the upcoming re-review of Participant 7's narrative:

1. Appropriate saliency of audio-visual information, including chat messages and audio alerts for gunfire and enemy missile launch detections.
2. Transparency of possibilities for action, including the amount of time needed for the Fire-X UAV to reach the ambushed Ground Troops, as well as the consequence of not launching any ALEs.
3. Accessibility of transitional state information, including the availability of information guiding a shift in priorities from clearing Route Jarus to supporting the Ground Troops under attack.

As a special note, in the following section the timeline narration of events (duplicated from Chapter 4) will be followed by a discussion of observability relevant to that passage of the narrative. Italicized font and text indentation will be used to separate the original narratives (pulled from Chapter 4) from the additional commentary that focuses on observability. More specifically, text that is indented and italicized indicates the participant's original detailed narration pulled from Chapter 4, while the text closer to the margins and un-italicized is the new expanded upon details of observability.

Participant 7's Detailed Narration – Observability

> *Participant 7 begins their course of action by having their "Fire-X" UAV and Ownship reconning Route Jarus, and their "Shadow" and "Gray Eagle" UAVs reconning NAI 26; these aircraft remain on these tasks for the first quarter of the 60-minute vignette. At 04:00 minutes into the vignette, Participant 7 receives a chat message with information from the first pop-up event reporting heavy enemy activity near the Ground Troops to the East of the dam. At 04:04 the participant positively identifies a group of hostile enemy armored personnel carriers (APCs) in NAI 26, which they*

> *choose to kinetically engage twice over the next seven minutes. Although the participant's attention was primarily focused on engaging the hostile APCs, mid-way through this engagement the participant briefly turns their attention over to read their chat message (at 06:25), 2 minutes and 25 seconds after receiving it.*

There was a lot of information that Participant 7 needed to collect in order to find, identify and engage the hostile APCs at minute 04:04. As previously described in the *TECUMSA: User Workflow* section, the typical process seen in these evaluations for confirming PID involved actively monitoring the map for entity icons, and the live video feeds of aircraft to get a line of sight on the potential threat. The PID process also often required the participant to actively direct (e.g., pan, zoom) the map and the live video feeds to better focus on specific entities. This process relied on an intimate coupling between perception and action because as the participant's search for information guided their activities, those activities subsequently guided the information they searched for.

Part of the challenge of observability is maintaining appropriate situation awareness in an information-rich environment such as this. As time goes on, information requirements change and the priority of tasks often shift. This was the case with the engagement of the hostile APCs. As Participant 7 was nearing the end of the engagement cycle of the hostile APCs at minute 10:00 by obtaining final visual confirmation of threat destruction through the live video feed, an enemy missile launch alert goes off (see Figure 21).

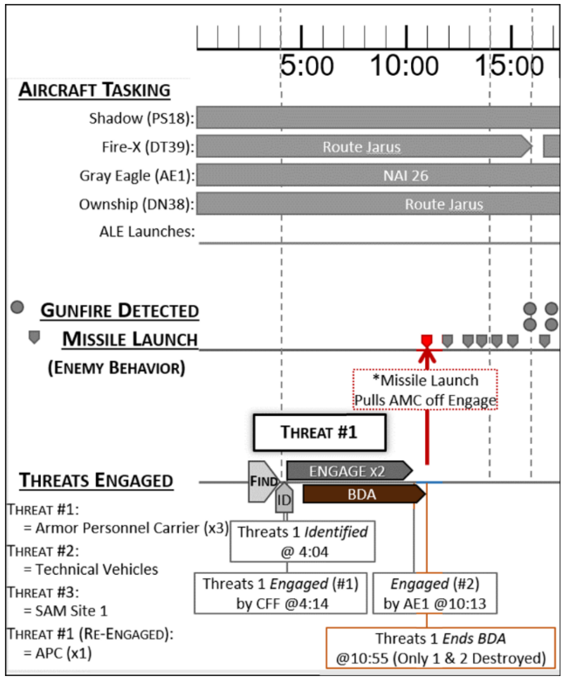


Figure 21. Highlighted Portion of Participant 7’s Event Timeline During an Enemy Missile Launch Alert (Vignette 4, Round 2 Evaluation).

This enemy missile launch alert occurs at minute 10:55, which was a high saliency alert in the interface, accompanied by multiple auditory warnings as well as large red icons in the map for each launch detection event. Prior to minute 10:55, no missile launches had occurred in the battlespace, which made the saliency of the audio alerts all the more attention-grabbing. The design of increased perceptual saliency of enemy missile launches (paired with participant’s real-world aviator training) meant that enemy missile launches were likely to take attentional precedence over other concurrent tasks; this was exactly what happened while the participant was confirming the APC threat destruction.

Once the participant heard the auditory alerts of “Missile Launch Detected,” they immediately dropped the engagement task and focused on the newest threat at hand, which was clearly firing at them or one of their UAV. Although the participant decided to prematurely wrap up their engagement of the APC, ultimately costing them time at the end of the vignette to re-find and re-engage it, the participant *was* able to quickly focus their efforts on the higher risk threat at hand (i.e., the enemy missile launch site that was actively firing missiles). The information presentation was therefore well-aligned to the goals/priorities of the participant; Participant 7's response time to the enemy missile launch was relatively quick and therefore allowed them to take decisive actions to manage safe flight for their aircraft near the enemy missile launch site (successful observability into the consequence of inaction). It should be noted that the autonomous route planner would only route around *known* counter-air threats. To be *known*, the operator had to mark the entity in the system as a type of counter-air threat, which would generate a "threat ring" around the enemy that was treated similar to an NFA by the route planner.

> *At 14:00 minutes into the vignette, the Ground Troops report via chat that an enemy patrol is approaching their position (the second pop-up event), which the participant reads 2 minutes and 6 seconds later. The previous two chat messages (at 04:00 and 14:00 minutes) were foreshadowing of the next chat message at 16:00, which is a critical turning point in the scripted mission priorities. At 16:00, the Ground Troops report via chat that they are now under attack (the official start of the third pop-up event, which is the most mission impactful) and request immediate armed assistance from the participant. Simultaneous to this incoming chat message, two audio/visual alerts of gunfire detection occur, along with a large orange icon in the map indicating the precise location of a gunfight that the Ground Troops are now in with an enemy patrol.*
>
> *Up until the time of the gunfire detection at 16:00, there have been 6 missiles launched at one or more of the participant's team of aircraft. Note the participant is not told which aircraft is getting shot at, only the grid location the enemy missile launched from. In the TECUMSA system, enemy missile launches were accompanied by both a salient auditory warning as well as a large icon in the map indicating the location of the*

> *enemy missile launch. Despite the saliency and time pressure of these missile launches, the audio/visual alert of the gunfire was prominent enough to grab Participant 7's attention. Shortly after noticing the gunfire, Participant 7 quickly tasks Fire-X to provide support to the Ground Troops.*

Figure 22 contains a portion of the Participant 7 timeline, deconstructed and repurposed to highlight the coupling of perception and action between minute 14:00 and minute 26:00, a timeframe that will be discussed in more detail in the following text. In Figure 22 the top row is *time* in minutes, the second row is *Participant 7's Task Priorities* at the time. The next row down contains the *Auditory Feedback* the participant received at that point in time, along with the frequency if they received it multiple times in the same minute (e.g., “x2”). The next row down contains the *Visual Feedback* the participant was focused primarily on at that point in time. Finally, the bottom three rows indicate when the participant *heard* the auditory feedback, *saw* the visual feedback, and *acted* on this information. Figure 22, therefore, depicts key pieces of information as well as meaningful actions that were taken as a result of seeing/hearing this information.

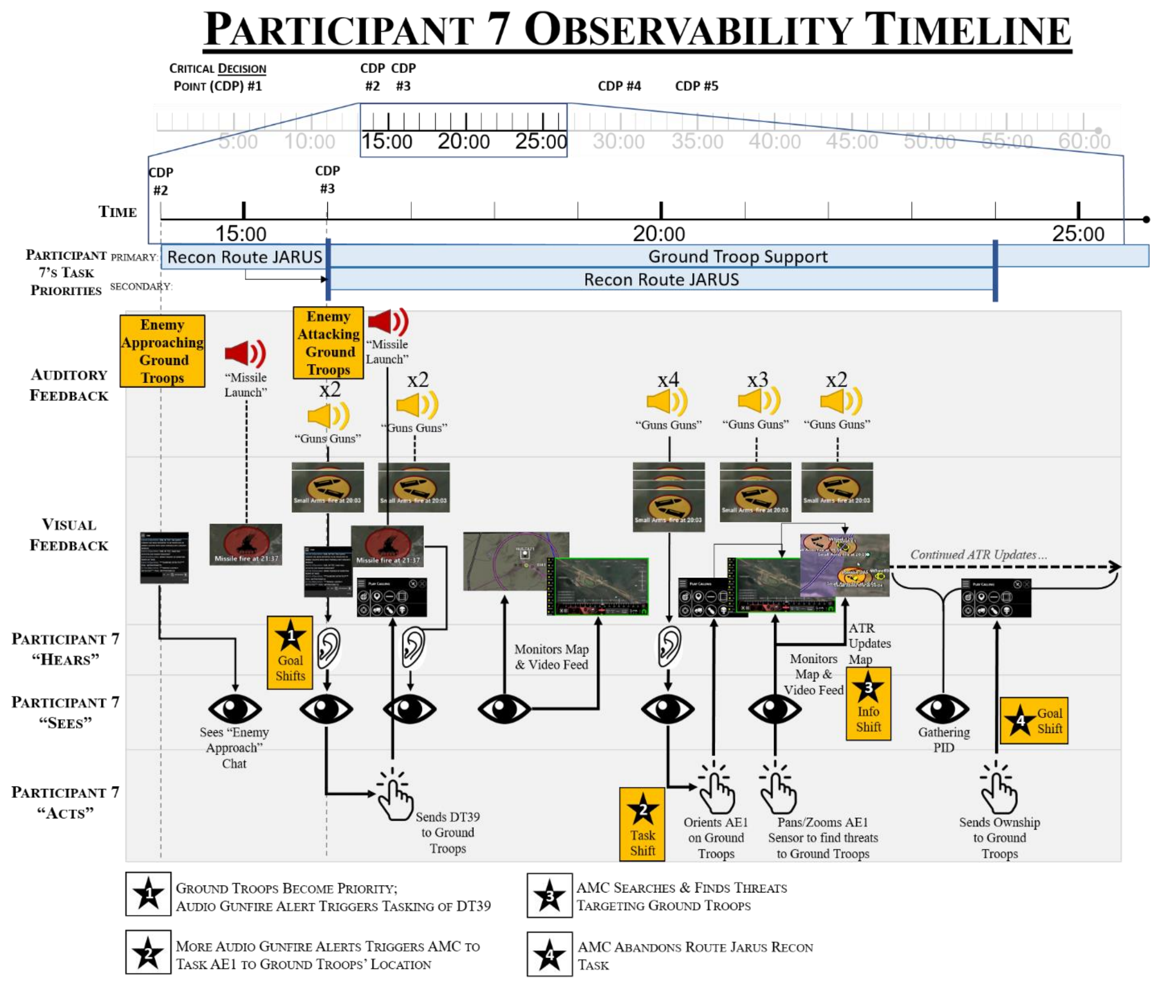


Figure 22. Deconstructed Timeline for Participant 7 from Minute 14:00 to 26:00 in Vignette 4, Round 2 Evaluation.

Just as the auditory saliency of the enemy missile launch detection at minute 10:55 had a significant impact on Participant 7's activities, so too did the auditory saliency of the gunfire detection at minute 16:00 in the vignette. More specifically, although Participant 7 actively monitored the chat tile throughout the vignette, there was often a delay between the chat message receipt and the message actually being read by the participant. Likely this delay was influenced by the permanency of the information, since participants could always go back to access an older chat message, typically without significant impact from the delay, but often could not easily tolerate a delay to the task at hand (e.g., tasking an aircraft, steering a camera sensor, confirming weapons release).

Unfortunately, the chat message sent at 16:00 minutes into the vignette was an exception to that rule and had significant consequences the longer it took for a participant to see the chat message since it contained the Ground Troop's request for immediate armed assistance (i.e., every second is potentially life-changing when troops are under attack). At the same time this critical chat message comes in, a gunfight breaks out in the scripted vignette scenario at the location of the Ground Troops, which triggers a separate audio-visual alert regarding Gunfire detection in the battlespace. As previously discussed, the gunfire alert is a high saliency alert in the interface accompanied by multiple auditory warnings as well as a large orange icon on the map. Prior to minute 16:00, no gunfire alerts had occurred in the battlespace yet in Vignette 4, which made the saliency of the Gunfire audio alerts all the more attention-grabbing.

Less than 30 seconds after hearing the gunfire alert, Participant 7 tasks the Fire-X (or "DT39") UAV to fly over to the location of the gunfire. Because of how quickly Participant 7 responded to the gunfire audio/visual alert, it's unclear whether they actually read the chat message or if they simply made the correct assumption regarding the cause of the gunfire alert (i.e., the gunfire involved the Ground Troops). In the case of the latter, it's completely reasonable that Participant 7 had the situation awareness to match the location of the gunfire alert icon in the map to the last known location of nearby friendly Ground Troops, to then infer that the Ground Troops needed assistance.

The reason for this in-depth discussion of the specific gunfire alert at minute 16:00 is because multiple participants only noticed the Ground Troops were under fire *because* of this audio alert. Additionally, the response time of participants to task aircraft to assist the Ground Troops was linked to their immediate perception of the auditory alerts of gunfire detection. Not only did the gunfire alert draw their attention to the fact that there was a new event occurring, but it also linked pertinent location information to the event (i.e., gunfire audio alert coincides with gunfire icon appearing in the map at the location of gunfire).

In interface design, there is always a balancing act to consider when choosing to make certain information high saliency. In *this* case, the gunfire detection alert was a pivotal moment to draw the user's attention to so they would know to help the Ground Troops under attack. However, the gunfire alerts may not *always* be the highest priority information to the user in an air assault operation. In other words, there were no false alarms (e.g., celebratory gunfire at a wedding) that the operator would be alerted to in this particular simulation. Designing alerts is tricky since the relative importance of information is contextually dependent and cannot always be known ahead of time by designers; different events carry a different meaning to the user at different times. In this synthetic task environment, by designing the gunfire alerts to be distinct from the enemy missile launch alerts it allowed the operators to maintain situational awareness over a range of high saliency audio feedback, thus successfully supporting user observability.

Although all three audio alerts were distinct from one another (i.e., new chat message, gunfire detected, and enemy missile launch detected), there is still the effect of desensitization or selective attention to certain alert types. In the case of the gunfire alert at minute 16:00, the saliency was likely enhanced by the fact that no other gunfire audio alerts had occurred yet (in Vignette 4) prior to minute 16:00. On the other end of this was the missile launch detections, which are typically of much higher importance than the gunfire alert (generally speaking), but had already occurred 6 times by minute 16:00.

> *At 20:41 the participant realizes how long it would take for Fire-X to reach the Ground Troops, and decides to also task the Gray Eagle to provide armed support to the Ground Troops under attack. Because the Gray Eagle was already in range of the Ground Troops, the participant was able to immediately start searching for entities in the location of the gunfire in hopes of finding the threats attacking the Ground Troops. After searching the map and video feeds for entities near the gunfire for 3 minutes and 27 seconds, the participant finds a few wheeled vehicles in the location of the gunfire but has not yet been able to positively identify them as friendly or enemy.*

A key question for TECUMSA interface designers after this evaluation was "Why is it that Participant 7 didn't know it was going to take Fire-X too long to reach the Ground Troops?" Part of the play calling process for the participant is to review the proposed route plan generated by the system. As is shown in Figure 23, this includes reviewing the map to see the recommended aircraft and proposed flight route generated by the system. However, unfortunately, there was no information for an estimated time enroute, nor the estimated time required to descend altitudes. Having observed the impact this information (or lack thereof) had on the participant's decision-making process in the Round 2 Evaluation, it would be a worthwhile addition to the play calling interface to present the aircraft's estimated time to reach an objective location. This change would provide valuable *predictability* of the projected future state of aircraft to enable participants to know *before* even starting the play whether or not the plan is suitable under their given time constraints.

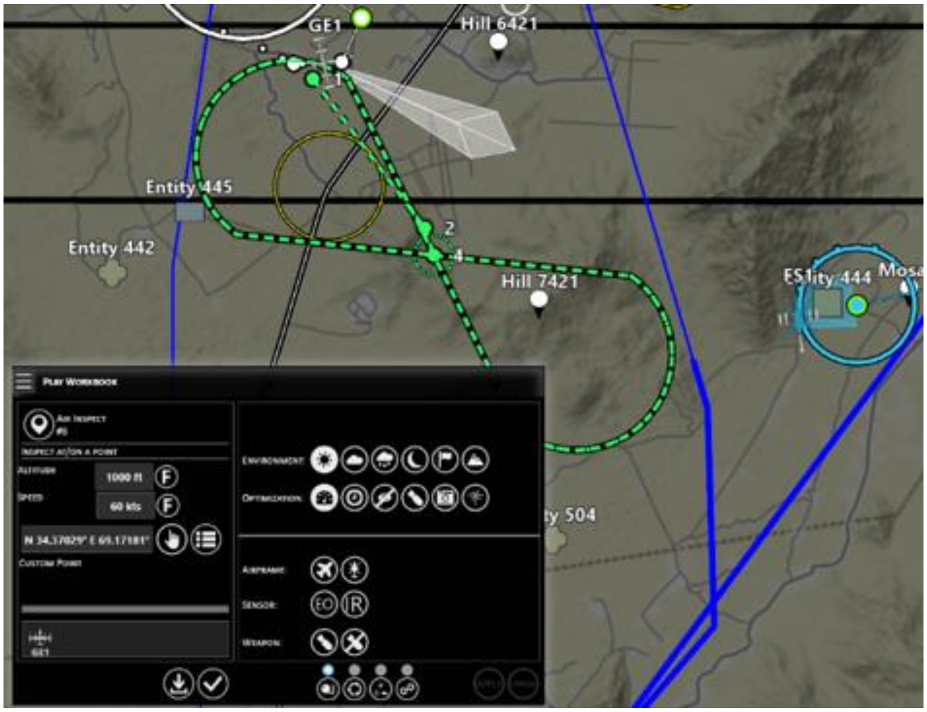


Figure 23. Screenshot of TECUMSA Interface just before Participant Approves the Play to be Executed.

Around 24:00 the participant notices that Fire-X is still several minutes away from the Ground Troops. Since Gray Eagle has only 1 missile remaining in its weapons payload, the participant shifts their attention to getting additional armed support over to assist the Ground Troops. At 24:08 the participant pulls the Ownship off the task of reconning Route Jarus and instead directs it to head over to support the Ground Troops, which becomes their third of four available aircraft (excluding the ALEs, which are not launched) tasked to assist the Ground Troops up to this point in the vignette. After tasking the Ownship, the participant returns to looking for the enemies engaging the Ground Troops through Gray Eagle's camera sensor.

At 26:20 the participant is able to positively identify the three hostile enemy technical vehicles attacking the Ground Troops, and then reclassifies the threats as "hostile" in the TECUMSA system. At 26:30 the participant calls an Engage Play using the Gray Eagle, and at 26:55 the participant launches missiles off of Gray Eagle at the hostile threats. At 27:20 the participant confirms the hostile technical vehicles are destroyed and has therefore successfully supported the unanticipated pop-up event of ambushed Ground Troops.

Figure 24 is a screen capture at roughly minute 25:00 in Participant 7's vignette, just before they see the threats to the Ground Troops and reclassify them as hostile. Note in Figure 24 the yellow clover icons in the map (indicating unknown entities) just to the east (right) of the orange gunfire detection icons, as well as Gray Eagle's camera sensor footprint projection in purple (indicating the sensor's visible field of view). As for the process of calling an engage play, it involves notable coordination between the human and the autonomous systems, including active monitoring of the Vehicle Dashboard to maintain PID and launch weapons once in range. Additional aspects of authority and role allocation (e.g., positioning aircraft, weapons release authority) will be discussed in the next chapter on directability.

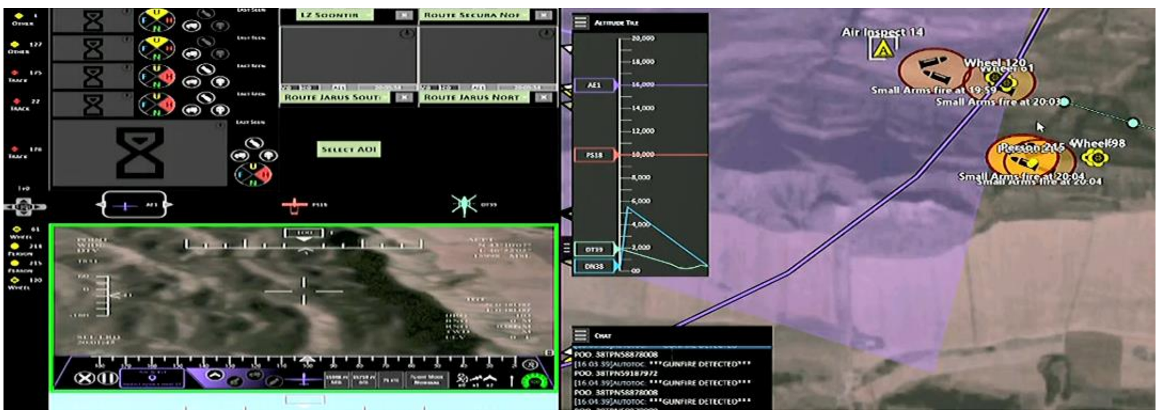


Figure 24. Participant 7's TECUMSA Interface just before Threats to Ground Troops are Located.

> *After neutralizing these threats, the participant returns to trying to precisely locate the Ground Troops location and communicates with the TOC (played by an experimenter) to get information on the status of the Ground Troops. The participant specifically asks the TOC over radio (via headsets) communication, "Are they [Ground Troops] alive?" Behind the scenes the TOC stalls to answer the participant's question, because in 55 seconds the fourth critical chat message is scheduled to be sent to the participant (at 30:00), requesting a medical evacuation (MEDEVAC) for the Ground Troops. In the meantime, as the participant waits for status*

information on the Ground Troops, they work to find threats from the surrounding areas of the Ground Troops' approximate location.

At 30:00 the fourth pop-up event initiates, where a chat message is sent to the participant requesting MEDEVAC support for the Ground Troops. At this point, the Ownship is still enroute to the Ground Troops. At 33:41 the participant checks their chat window and sees the Ground Troops' request for a MEDEVAC. By the time the Ownship reaches the Ground Troops location, at 34:01, four missiles have been launched at one or more of the participant's aircraft. It is at this arrival time at 34:01 that the experimenters notice the Ownship aircraft hovering just above the ground and barely moving forward. This was because, upon arrival to the Ground Troops, Participant 7 had the pilot position the Ownship to land and actually carry out the MEDEVAC. It is worth noting that no other participant did this, and behind the scenes experimenters were extremely impressed by the participant's dedication to help the injured Ground Troops.

After the participant offers to provide the Ground Troop's MEDEVAC, 59 seconds go by. The fifth and final pop-up event, which is a scheduled chat message, is then sent at 35:00 stating that a MEDEVAC aircraft team is 25 minutes from the current Ground Troop's location and will fly ingress via Route Secura and land in LZ Soontir. Since Participant 7 has their Ownship positioned to actually carry out the MEDEVAC, the TOC adapts the script by further clarifying to the participant over radio communications to "Charlie Mike" (or Continue Mission) because the Ground Troops need specialized medical equipment.

At 35:00 the participant turns their efforts towards clearing the surrounding area of threats. After 4 minutes and 36 seconds of searching, the participant positively identifies an anti-aircraft missile system (hereon referred to as a "hostile missile system"), which has been launching several missiles at the participant's aircraft throughout the mission. At 39:58 the participant sends the first LRPF at this threat (i.e., calling for a missile launch from a remote ground location), and after not observing the desired kinetic effects, the participant decides at 48:20 to send a second and final LRPF at the hostile missile system. Shortly after the second LRPF, the participant confirms the hostile missile system is destroyed through a nearby aircraft video feed and marks it as destroyed in the TECUMSA map.

> *At 49:00 into the vignette, 14 minutes after being notified of the new route and landing zone needing to be reconned (i.e., Route Secura and LZ Soontir) for the Ground Troop MEDEVAC, the participant still has three of the four aircraft loitering nearby to the Ground Troops. In other words, the participant has not tasked any of the aircraft to specifically search along Route Secura or LZ Soontir. At 56:16 the participant finds one of the APCs in NAI 26 that had not previously been impacted by the first engagement. At 56:19 the participant sends an LRPF request, and at 56:40 confirms the APC threat has been destroyed.*

One of the more noteworthy aspects of Participant 7's Vignette 4 run is the fact that they never launched any of their available ALEs. In this vignette, all participants were provided four available ALEs equipped with camera sensors, that they could launch off of the Grey Eagle or the Ownship (two ALEs were onboard each aircraft). Although ALEs were not the fastest flying aircraft, having four additional assets to recon the battlespace would have surely helped Participant 7 to search for threats along the new Route Secura and LZ Soontir (third pop-up event). One of the lessons designers took away from this is that perhaps the pros and cons of launching the ALEs were not represented in a meaningful enough way to participants. Future development efforts of TECUMSA will consider including *unlaunched* aircraft as part of TECUMSA's autonomous reasoning about optimal aircraft tasking solutions, so additional tasking possibilities can be recommended to the user.

> *At 60:00, the end of the vignette, the TOC asks the participant whether or not they advise the MEDEVAC operation to continue. Since the participant had neutralized the known hostile missile system and did not detect or know of any other threats that would impact the incoming MEDEVAC team, they gave the TOC an affirmative "Go" call to continue with the MEDEVAC. After consulting with SMEs after the completion of this run, both SMEs observing the vignette agreed that this was a good call made by the participant.*

Participant 7 Summary

Observability in interface design means determining if the user interface provides adequate information for the user to see the possibilities and consequences of action relative to their goals. In Participant 7's Vignette 4 (Round 2) simulated air assault mission, there emerged several events where the user's interactions with the system either did or did not meet designer expectations. However, rather than look at these events to say whether the information representations were *right or wrong, it is better to say the representations were well-aligned or misaligned* to support Participant 7's tuning to affordances in this particular context. This dichotomy will be used to recap the various noteworthy events from Participant 7's Vignette 4 (Round 2) run.

Well-Aligned Observability – Participant 7. To begin, there were information representations in the TECUMSA interface that were *well-aligned* to support Participant 7's tuning to affordances. One example of this designed observability was the accessibility of *transitional state information*, specifically when there was a necessary shift in priorities (mid-mission) from clearing Route Jarus to supporting the Ground Troops that were under attack. In the Round 2 interface design, the saliency of the audio-visual alerts for gunfire detections was able to correctly direct Participant 7's attention to the information that signaled the Ground Troops were being ambushed.

One of the key ways participants were able to maintain situational awareness during the simulated air assault missions was through the audio-visual alerts TECUMSA provided. In the Round 2 Evaluation, some of the more significant shifts in the state of the battle were accompanied by chat messages, gunfire alerts, or enemy missile launch detection alerts. The importance of these three events thus warranted the addition of auditory feedback in addition to visual feedback. While the pairing of audio-visual feedback was often salient enough to draw the participant's attention during the simulated mission, some refinements to the design are still possible. One design refinement could be to the chat messages, which had varying levels of priority, but did not have varying levels of audio-visual feedback. A simple ability to mark a chat message as "urgent" by the sender, so the interface could augment increased saliency of audio-visual feedback, would likely prove useful in a wider range of real-world circumstances.

Misaligned Observability – Participant 7. There are also several events where the information representation was misaligned and did not sufficiently support Participant 7's tuning to affordances. One example arose when Participant 7 chose to send Fire-X over to support the Ground Troops partway through the mission (per the autonomy's recommendation). The participant's lack of experience with these UAVs meant they were not able to establish mental models for typical transit times of aircraft. This meant that Participant 7 found out too late that Fire-X was not going to reach the troops in an acceptable amount of time, and another aircraft would need to be sent to support the troops. This revealed a tuning misalignment in resource affordances, and more specifically that flight time information needed to be represented in the interface ideally during the play plan reviewing process. This event highlighted a theme for interface design refinements, which is to have the *appropriate information at the point of need*. In this case, flight times needed to be given within the play calling interface, so the participant could evaluate aircraft transit times *before* confirming the execution of a play.

Another case involved Participant 7 not clearly seeing the consequence of not launching any of their four available ALEs, which ultimately meant that only 3 of their

11 reconnaissance tasks would be addressed. Participant 7's aircraft tasking strategy might have been different if they had a more holistic view of the sensor coverage they were able to achieve throughout the mission. One idea would be a heat-map layer in the map, which shows the recency and/or duration of sensor coverage throughout the battlespace. This type of display could have highlighted to the participant the concentration of forces at the site of the Ground Troops, and relative sparsity of sensor coverage in other areas of interest in the battlespace. Alternatively, the task of achieving sensor coverage of the battlespace could be offloaded onto the autonomous systems of TECUMSA. More specifically, allowing the system to autonomously task idle aircraft and/or idle sensors could have greatly increased the number of reconnaissance tasks that were accomplished.

Discussion – Participant 7 Observability. One noteworthy discussion of designed observability in Participant 7's Vignette 4 (Round 2) was the audio-visual alerts for gunfire detections and enemy missile launches, which were well-aligned for *this context* (i.e., synthetic task environment). However, these audio-visual representations may not be equally effective across all task environments. More specifically, one element that perhaps made the audio alerts more effective was the absence of *false alarms*.

Additionally, in real operational environments there is typically much more competing audio information that is transmitted inside the AMC's cockpit/headset, and not all of it is always actionable or contextually relevant. Additionally, while the audio-visual alerts regularly captured all seven of the participants' attention in the Round 2 Evaluation, there was still the need for additional meaning interpretation. For example, one participant misinterpreted the gunfire as coming from *only* hostile sources and did not realize for quite some time that friendly Ground Troops were actually involved in the gunfight as well.

Observability Chapter Summary

There is a reason user interface design is considered an iterative process; a design that works well in one context may be insufficient in another (and vice versa). In fact, a design that works well for one person may not be ideal for another (e.g., differing tactical experience, differing interface familiarity). The two rounds of formal evaluations of TECUMSA in a synthetic task environment provided numerous opportunities for interface designers to evaluate how well the chosen representations of information enabled participants to see the solution space in terms of the possibilities and consequences of action (i.e., observability).

When observed user behaviors misalign to the expectation of behavior, it helps interface designers to build out theories that compare the user's *perceived* consequences of actions to the *actual* consequences of action. Although only Participant 7 and Participant 2 are discussed in detail, the decomposition of their Vignette 4 run combined with observations from both the Round 1 and Round 2 Evaluations revealed three key themes of designing observability emerged: *Providing appropriate information at the point of need, managing cognitive resources,* and *representing boundaries and transitional states.*

The first theme for perception-action coupling is *providing appropriate information at the point of need,* which involved providing the necessary information to a user, in a meaningful format, and at the time and location of need. The second theme is *managing cognitive resources,* which covers both the cognitive and physical resources required for a user to gather information and make system inputs. The third and final theme is *representing boundaries and transitional states*, which involves clearly representing constraints/consequences on action, particularly as priorities shift over time, as well as violations to "normal functioning" within the system (e.g., aircraft running out of fuel).

For some mission events the design was well-aligned to the pertinent affordances of the task (e.g., saliency of gunfire alerts signaled users to transitional priority states), and for other events the design was misaligned to the user's informational needs (e.g., unclear temporal constraints of different aircraft to move across the battlespace, consequences of insufficient sensor coverage throughout the battlespace). These events help to expand the interface designer's understanding of the operational information requirements, which will help ground future efforts to further refine the interface design. Regardless of whether the interface was well-aligned or misaligned to task affordances *in this mission context*, the complexity and unique courses of action permitted in this synthetic task environment revealed several opportunities to further refine observability in the TECUMSA interface across a wider range of use cases.

Observability in Design for Participant 2

Four major elements of observability will be discussed in the upcoming re-review of Participant 2's narrative:

1. Providing information representations that are useable. More specifically, discussing the consequences of having the chat tile too small.
2. Managing cognitive/physical resources. More specifically, Participant 2's situation awareness and battlespace monitoring strategy allowing them to quickly

discover that the Shadow UAV was an inefficient aircraft to task due to the required transit time.

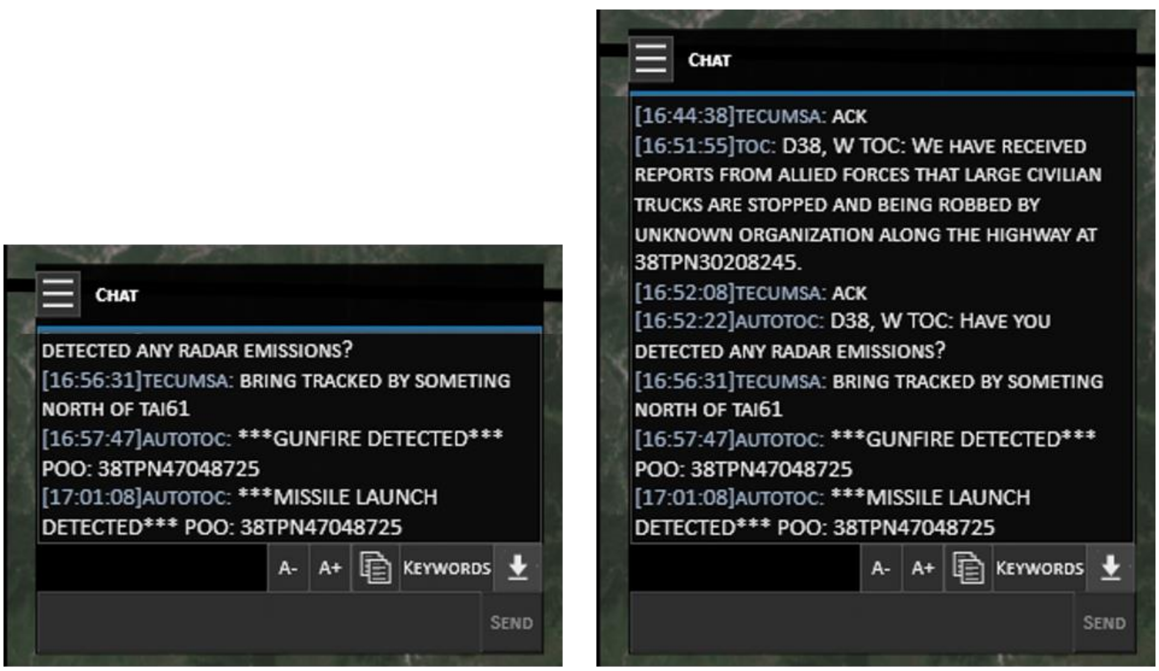
3. Representing boundary conditions/transitional states. Specifically, Participant 2 experienced a significant shift in consequences when their Fire-X UAV traveled in range of an un-templated threat and was shot down. The loss of this aircraft created a different set of action possibilities that the participant had to then adapt to (i.e., required redistribution of tasks due to the loss of an armed aircraft on the team).
4. Providing appropriate information at the point of need. In this case, Participant 2's misinterpretation of the color faded icons (indicating the last *known* location) meant unnecessary use of weapons on an entity that was no longer at its last seen location.

Recall, text that is indented and italicized indicates the participant's original detailed narration pulled from Chapter 4, while the text closer to the margins and un-italicized is the new expanded upon details of observability.

Participant 2's Detailed Narration - Observability

> *In the first five minutes of Vignette 4, Participant 2 has their "Fire-X" UAV tasked to recon NAI 12, their "Gray Eagle" UAV reconning TAI 63, and their "Shadow" UAV and Ownship tasked to do a joint recon of Route Jarus. At 04:00 minutes into the vignette, Participant 2 receives a chat message pertaining to the first pop-up event reporting heavy enemy activity near the Ground Troops to the East of the dam, but the participant never read this chat message. At 06:29, the participant pulls the Shadow off of the joint recon of Route Jarus after noticing that it will take a significant amount of time for the Shadow just to travel across the battlespace and start reconning Route Jarus. The participant then instead tasks the Shadow to recon NAI 11, which was a much closer area to the starting location of the Shadow.*

A significant contributing factor as to why Participant 2 might not have seen all of the received chat messages is the *size of their chat tile*. It was up to participants to increase the size of their chat tile from the default size, which had 7-lines of text visible without scrolling (see Figure 25 for comparison of Participant 2 and Participant 7's chat tile). Since Participant 2 did not resize this tile, chat messages quickly became buried in the hidden thread of messages. This occlusion issue was amplified by the fact that both chat messages *and* notification messages populated into the same chat thread, which quickly filled the 7-lines of visible text with notifications (e.g., gunfire detection, enemy missile launch detection), quickly burying higher importance chat messages. It should be noted that participants were trained on how to resize tiles in the interface, but few participants used this customization option.



(a) (b)

Figure 25. Vignette 4 Chat Tiles: (a) Participant 2's Chat Tile with 7-Lines of Text in Viewing Window, and (b) Participant 7's Chat Tile with 15-Lines of Text in Viewing Window.

Participant 2 noticing at 06:29 that the Shadow UAV would take too long to travel across the battlespace is an interesting point of contrast to what happened with Participant 7. Unlike Participant 7, Participant 2 noticed just a few minutes after assigning the task that the aircraft would take too long to fly to its destination. Although both participants dealt with the same lack of transparency into transit times, it appears Participant 2's constant monitoring of individual aircraft helped them to catch necessary adjustments sooner in the mission progress. This finding also reflects conversations with experienced aircraft pilots on the team. SMEs on the team referenced a trained technique of using a constant search/scan pattern to maintain situation awareness across a variety of cockpit displays to prevent "tunnel vision" or fixation on any one item for too long.

> *At 06:42 the participant launches the first ALE off of their Ownship, and then at 06:51 launches the second ALE off of the Ownship. With the two available ALEs on the Ownship launched, the participant then tasks both aircraft to inspect the location of a templated hostile missile system. Once the two ALEs are tasked, the participant turns their focus to finding hostile threats in the battlespace.*

It is worth noting here that again, the issue of transit times for aircraft was an observability misalignment. For the entire vignette these two ALEs were "enroute" somewhere, but never actually reach their destination. It would have been very beneficial for the users to understand how long it will take the ALEs to reach their destination *before* launching them, especially since the aircraft carrying the ALEs was faster flying and could have perhaps gotten the launch point closer to the destination point to help optimize on task completion times.

> *At 13:49 the participant identifies a wheeled vehicle near the vicinity of the templated hostile missile system. At 14:00 minutes into the vignette, the Ground Troops report via chat that an enemy patrol is approaching their position (the second pop-up event), which the participant*

> *never reads. After confirming no friendlies are in the nearby location to the templated hostile missile system, the participant calls an LRPF on a wheeled threat at that location at 15:06. At the exact same time (15:06), a gunfire detection audio/visual alert is generated. Five seconds later at 15:11 the third critical chat message is sent, reporting that the Ground Troops are now under attack and request immediate armed assistance from the participant (the official start of the third pop-up event, which is the most mission impactful).*

To identify the hostile affiliation of the wheeled vehicle near the templated hostile missile system, Participant 2 utilized additional information to the typical PID process described in the *TECUMSA: User Workflow* section. More specifically, the participant used the presence of simulated Radar Warning Receiver (RWR) rays intersecting at a nearby location to the hostile missile system (see red and orange lines in Figure 26). Participant 2's use of the RWR rays was not a unique strategy, but worth mentioning as it validated the intended use of this design feature, as it specified the higher-order relationship between hostile threat locations and the converging RWR rays.

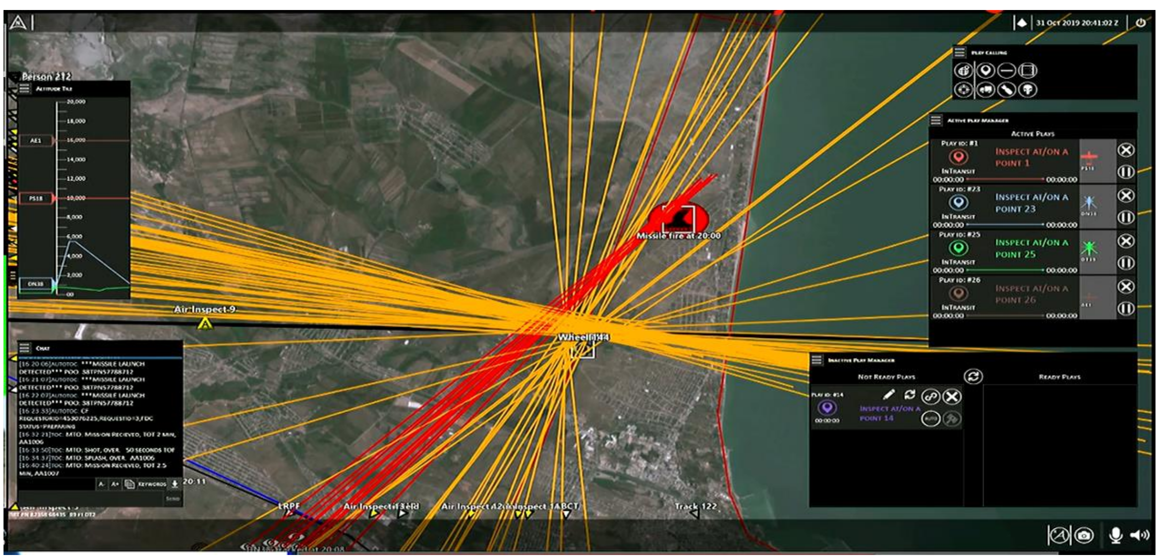


Figure 26. Map Design of Higher-Order Relationship between Hostile Threat Locations and the Converging RWR Rays (Red and Orange Lines).

> *Although the participant does not immediately read the third chat message reporting the Ground Troops under attack, the audio/visual alert*

> *of gunfire at 15:06 was salient enough to draw their attention to the Ground Troops location on the map. Up until this point, the participant has not read any of the critical chat messages tied to the scripted pop-up events in Vignette 4. However, Participant 2 correctly decides that the gunfire is potentially coming from and/or directed towards the Ground Troops that were known to be roughly in the area of the gunfire, based on their pre-mission briefing.*

Just like Participant 7, Participant 2's attention was drawn to the ambushed Ground Troops because of the salient audio from the gunfire detection alerts paired with the gunfire icon that populated the map at the location of the gunfire (see the *TECUMSA: Audio-Visual Feedback* section for details).

> *At 16:13 the participant has the Ownship head towards the gunfire, but at a safe distance to the South. After redirecting the Ownship and pulling it off of the Route Jarus reconnaissance, the participant checks their chat tile for additional information in regards to the gunfire detection. At 17:10 the participant reads that the Ground Troops are requesting immediate armed support from the participant. At 17:24 the participant sees multiple wheeled vehicle icons pop up on the map at the location of the gunfire. Participant 2 then spends the next 38 seconds confirming that the newly detected wheeled entities are indeed hostile and not the Ground Troops fighting back.*

The size of Participant 2's Chat Tile creates a challenge of observability since scrolling would be required to view messages outside of the 7-line max viewing window (see Figure 25 for chat tile reference). In this case, the gunfire alerts each took 2-lines of text and populated the chat *after* the message was sent requesting immediate armed support for the ambushed Ground Troops. This meant that without scrolling, the participant could not see these critical chat messages tied to the second scripted pop-up event, as it was buried by alert notifications. The result of this small Chat Tile was not only a delay in responsiveness, since the participant had to figure out there was missing information, but also extra effort to search within the chat to find the information they needed from the Ground Troops.

At 16:13 the participant tasks the Ownship to head towards the gunfire and assist the Ground Troops, and at 17:24 the participant sees multiple wheeled vehicle icons. Because the Ownship was already nearby the gunfire detection site, there is a relatively short time between tasking the Ownship and getting eyes on the wheeled vehicles in the area. Although this was an efficient redistribution of tasks, since the Ownship was armed and already in the vicinity of the Ground Troops, it would have been more optimal had an additional aircraft been tasked to finish the task of reconning Route Jarus. This sort of sensor coverage optimization is a powerful role for the autonomous systems to take on in future iterations of TECUMSA, where the autonomy is given the authority to task idle aircraft/sensors.

> *From 18:02 to 20:43 the participant is directing the Ownship pilot to get in position to engage the hostile threats attacking the Ground Troops. From 20:43 to 23:04 the participant is confirming the location of the Ground Troops to make sure they are not in close proximity to the intended kinetic strike location. At 24:10 the participant calls an Engage Play using the Ownship, and at 24:30 the participant launches missiles off of their Ownship at the hostile threats. At 25:19 the participant confirms the hostile technical vehicles are destroyed and has therefore successfully supported the unanticipated pop-up event of ambushed Ground Troops.*

The process of Participant 2 positioning the Ownship aircraft for the engagement, and confirming the location of friendly Ground Troops, showcased the decision-making of an experienced pilot trying to avoid any accidental friendly fire. These activities are important and sophisticated behaviors of human operators. However, in this synthetic task environment there was not a GPS-enabled network simulating Blue Force Tracking (BFT). Without BFT, Participant 2 had a hampered ability to determine the safe course of action without first knowing where exactly the Ground Troops were. This meant that the participant had to cross-check various information sources to verify the location of the friendly Ground Troops, including visually inspecting the Vehicle Dashboard video feed, verbal exchanges with the TOC, and referencing geographic locations mentioned in the chat messages. In future designs, the user should ideally be able to easily observe the consequences of a planned kinetic strike.

> *Following this engagement, at 25:53 a missile is launched at one of the participant's aircraft. At 26:25 the participant directs their attention to the location of the missile launch site. At 27:17 the participant identifies the missile launch site as hostile and at 27:27 sends a request for LRPF at the location of the enemy missile launch. From 25:53 to 27:49, four missiles in total are launched at the participant's aircraft. At 29:12 the participant learns that the Fire-X UAV aircraft has been shot down by an enemy missile (i.e., one of the enemy missile launches hit its target). At 30:00 the fourth pop-up event initiates, where a chat message is sent to the participant requesting MEDEVAC support for the Ground Troops.*

One of the most valuable attributes of this synthetic task environment was the structured events situated within a dynamic scenario. Just like no two missions in the real-world are ever exactly the same, slight adjustments in an operator's courses of action could produce very different outcomes. In this case, Participant 2 ended up losing one of their UAVs roughly halfway into the mission, due to the Fire-X's proximity to an unmarked enemy missile launch site. It's important to note that two key pieces of information are required to properly address a missile launch: A) Knowing where the hostile missile launch came from, and B) Knowing what aircraft are in dangerous proximity to the missile launch site.

In the Round 2 design of TECUMSA, the first major challenge was *knowing where the hostile missile launch came from*. One particular challenge participants had when trying to determine where an enemy missile launch happened in the map, occurred when multiple missiles were launched from the same location (which occurred the majority of the time). The issue was with discernibility, because of the occlusions of one missile launch icon layering on top of one another in the map (see Figure 27).

As Figure 27 shows, the most visually salient change in the map occurs on the first enemy missile launch. After the first missile launch however, it becomes very difficult to visually discern that additional missiles have been launched, without looking at the timestamp. Although the use of semi-opaque color fill in the icons was a somewhat perceivable difference when only two missile launch icons were stacked, it became exponentially more difficult to discern as more icons were stacked on top of one another (moving left to right in Figure 27). To add to the difficulty in observability, there could be more than one missile launch location, and the eye-catching motion of an icon “popping up” in the map was missed if the operator did not have that part of the map visible, for example, if they had the map zoomed in or panned to a different area in the map.

Figure 27. Layered Enemy Missile Launch Icons in Map at an Increasing Frequency of Occurrence.

The second major challenge was *knowing what aircraft(s) are in dangerous proximity to the missile launch site*. One notable challenge for the operator’s observability of aircraft locations was the visual clutter in the map (i.e., clustering or overlapping of information in the map). Although the participant could use their situational awareness to know generally where each aircraft was located, they could not solely rely on their mental model of the battlespace to know when exactly an aircraft was in range of an enemy missile launch site. Because of this, operators had to utilize the map to see current aircraft locations, and the visual clutter (e.g., entity icons, flight paths, areas of interest) likely made it difficult to quickly see where the Fire-X was in relation to the missile launches. Furthermore, recall that the autonomous route planner would only route around *known* threat rings if the participant had marked the entity as a type of counter-air threat in the system (which would be treated similar to an NFA by the route planner).

> *With Fire-X no longer available for tasking, the participant returns to their task of finding threats in the battlespace. From 31:58 to 37:50 the participant focuses on finding threats in TAI 63. At 37:50 the participant reorients their attention back to templated hostile missile system. At 38:51 they focus specifically on the icon in the map for Wheeled Entity 152, which is in proximity of the templated hostile missile system. However, in actuality, the Wheeled Entity 152 has moved from that location, and the color faded icon in the map only indicates the last known location. The participant misinterprets what the icon in the map indicates and decides to pursue an engagement of the Wheeled Vehicle 152 due to its proximity to the templated threat (i.e., the hostile missile system). At 41:00 the participant calls an Engage Play on Wheeled Entity 152 using the Gray Eagle.*

Participant 2's misuse of the color faded entity icons is worth noting because of the consequences of this misinterpretation. In the TECUMSA interface, the color faded entity icons indicated that the entity was no longer seen by any of the camera sensors, and the color would fade gradually more grey over the course of a few minutes to indicate if the entity was recently lost from view, or had been lost from view for several minutes. This interface feature was exceptionally helpful for participants in both the Round 1 and

Round 2 interface design, as it helped the operator maintain awareness of recent entity activities while reducing the saliency of older entity activity information. However, the semantics behind what the color faded icons indicated was implicit information that users had to learn, which unfortunately in Participant 2’s case was not always properly interpreted.

Figure 28. Color Faded Entity Icons in TECUMSA the Map.

> *At 42:47 the fifth and final pop-up event initiates, where a chat message is sent to the participant stating that the MEDEVAC team is 25 minutes from the current Ground Troop’s location and will ingress via Route Secura and land in LZ Soontir. Eight seconds after receiving the chat message the participant reads it, and through the think-out-loud procedure states that they need to locate LZ Soontir. However, before shifting their attention to LZ Soontir, the participant checks back in with Gray Eagle’s progress on the engagement of Wheeled Entity 152. The participant monitors and controls Gray Eagle’s camera sensor until 44:30, at which time they then decide to pivot the Shadow’s efforts from NAI 11 to instead begin addressing the need to recon LZ Soontir.*
>
> *At 44:35 the user interface begins to lag in response to the participant’s inputs, due to unforeseen artifacts of the synthetic world environment. Although verbally stating their task priorities, the participant*

struggles to work around the delayed system feedback. The system lag continues to be a challenge for the participant, and at 48:27 the experimenters choose to end the vignette. Thankfully this run through Vignette 4 was sufficiently long enough to capture the participant's chosen course of action and adaptation strategies to all of the scripted pop-up events.

At 48:27, the end of the vignette, the TOC asks the participant whether or not they advise the MEDEVAC operation to continue. Since the participant had neutralized the known threats in the battlespace, they confirmed with the TOC that the incoming MEDEVAC team could continue their mission, but advised of a known radar detection system to the North of the Ground Troops. After consulting with SMEs after the completion of this run, both SMEs observing the vignette agreed that this was a good call made by the participant.

Participant 2 Summary

Just like with Participant 7's, in Participant 2's Vignette 4 (Round 2) simulated air assault mission there emerged several events where information was either well-aligned or misaligned for them to see the possibilities and consequences of action relative to their goals. The following is therefore a review of the various noteworthy events from Participant 2's run.

Well-Aligned Observability – Participant 2. To begin, there were information representations in the TECUMSA interface that were *well-aligned* to support Participant 2's tuning to affordances. One example of this was Participant 2 being properly tuned to the possibilities for action with the Shadow UAV. More specifically, they recognized the limitations on travel time for the Shadow, which allowed them to quickly re-task the aircraft to a closer objective rather than wait for the aircraft to reach a further task location. There are likely multiple reasons why this participant was able to properly tune to the affordances of the Shadow including, their management of cognitive resources by employing frequent visual scans of the battlespace to maintain up-to-date information, as well as their experience with aircraft with different max airspeeds. The primary lesson learned from this observation was that increased transparency into transit times needs to be provided *before* the play is actually started by participants, to avoid unnecessary trial-by-error.

Misaligned Observability – Participant 2. There were also several events where the information representation was misaligned and did not sufficiently support Participant 2's tuning to affordances. One noteworthy feature of the user interface was the customizable chat tile size. However, in the evaluation resizing the chat tile from the default size (with 7-lines of text in the viewing window) was the responsibility of the Participant. Because Participant 2 did not resize their chat tile, it meant that they ultimately had to gather information in the chat by scrolling and searching for messages. Future design improvements for this include giving participants more time for pre-mission configuration customization so the user can resize tiles as needed, separating chat messages from notification messages to avoid burying high importance information, and including the ability to mark a chat message or sender as higher urgency to increase audio-visual saliency.

Another example of misaligned observability in design was in the representation of boundary conditions/transitional states, and providing appropriate information at the point of need. Specifically, Participant 2 had their Fire-X aircraft fly into range of a hidden threat, which then subsequently fired at and ultimately shot down the Fire-X UAV. This event highlights two important pieces of necessary information: A) Knowing where threats and hostile missile launch sites are located, and B) Knowing what aircraft are in dangerous proximity to the threats and enemy missile launch sites. The user interface is therefore very important to providing the user with persistent situation awareness in regards to threats and aircraft proximity to threats. In future designs, ideally once an enemy missile launch is detected a threat ring would automatically drop around the area to offload the risk management task onto the autonomous route planner (rather than require the user to mark the threat before the threat ring pops up).

Additionally, the loss of the Fire-X UAV aircraft created a transitional state, or different set of action possibilities that Participant 2 had to then adapt to, including a redistribution of tasks due to the loss of one armed and sensor equipped aircraft on the team. Thankfully the participant was quickly made aware of the loss of this aircraft and was also able to quickly adapt their course of action to the change in action possibilities. However, this is another opportunity to increase the number of tasks allocated to the autonomous systems, as this could be a strength of autonomy to do a quick redistribution of ongoing plays if one aircraft is shot down or pulled off the team.

In line with the theme of providing appropriate information at the point of need was Participant 2's misinterpretation of the color faded icons. The color faded entity icons in the map represented the *last known location* of an entity, whereas the vibrant color indicated *it is currently at that location* (i.e., currently in sensor view). The cost of this misinterpretation for Participant 2 was the amount of time/resources they spent trying to find and engage an entity that was no longer at that location. Unfortunately, the limited training time on the TECUMSA system meant that participants had a lot of information to manage, which only added cognitive workload to an already cognitively demanding mission. As such, future evaluations could potentially allow for additional learning time on the TECUMSA system, to help increase the accuracy and automaticity of information processing with a new interface.

Discussion - Participant 2 Observability. Participant 2 had a theme of strong situational awareness. From their quick catch that Shadow was not going to reach its destination soon enough, to the correct assumption that the gunfight involved friendly Ground Troops based on the location of gunfire detection on the map. However, managing a team comprised of this many aircraft, in a contested battlespace, is not a simple task (e.g., maintaining situation awareness, choosing where to allocate resources).

Two aspects of designing for a diverse set of users, therefore, emerged from these evaluations and Participant 2 specifically: A) Reduce the need to "interpret" information by striving to have meaning/affordances directly perceivable. In other words, reducing the need for mediating knowledge not only decreases cognitive effort but also reduces the risk of information being incorrectly interpreted (e.g., misinterpreting what a faded entity icon means, or what the gunfire detection alert actually indicates). B) Free up cognitive bandwidth of users by offloading lower risk decision making to automated systems. For example, allowing automated aircraft tasking when an aircraft or its sensor payload is idle, rather than requiring the human operator to task the aircraft. Task allocation across both human and autonomous agents also has the potential to better utilize the unique expertise of the human operator, rather than consume the human's time/effort/attentional resources on low-risk trivial tasks.

Designers must be aware of information that is open for interpretation. For example, one participant misinterpreted that the gunfire alert was coming from *only* hostile sources, and did not realize for quite some time that friendly Ground Troops were actually involved in the gunfight as well. Thankfully, necessary friction was put into the design to help prevent human error if the *wrong* interpretation was to be made, as the HMI required participants to mark an entity as “Hostile” before they could kinetically engage it (a feature of directability that will be discussed more in Chapter 6). However, this is not necessarily an infallible design. The extra step in the engagement process, which forces a user to first reclassify the entity they want to engage as “hostile, at least forces participants to make a conscious decision about the identity of the entity before firing any weapons at it. In the future, more advanced HMI features could be explored. For example, having a set of images pop-up of the entity before the user engages so they can get a clearer visual confirmation of the possible threat, or perhaps a warning from the system if the ATR disagrees with the user’s choice to engage a particular entity type.

CHAPTER 6 – Directability

Introduction

As stated in earlier chapters, the goal for the HMI of TECUMSA was to provide the operator/AMC with the necessary situation awareness and interface usability to enable the execution of an air assault mission, including command and control of a team of UAVs and manned aircraft. The focus of this chapter is therefore on the elements of the TECUMSA interface that afforded directability of behaviors, including the user's ability to direct behaviors of an automated partner's resources, activities, and priorities (McDermott et al., 2018). Also within the scope of this chapter is designing directability for the user's ability to take action, such as having the authority to fire weapons at enemy threats or manually steer a camera sensor to view a location of interest.

When designing for humans to effectively coordinate and collaborate with autonomous agents, it may help to further distinguish observability from directability. Simply put, *observability* is about the human having the ability to understand and explore the solutions generated by autonomy. *Directability* is about the autonomy knowing what meaningful solutions to generate in the first place, which the user can guide, shape, and adjust as needed. In other words, observability is about *seeing* the solution space, whereas directability is about the *authority/ability to select* the appropriate solution.

One of the key ways directability is achieved in the TECUMSA system is through an approach referred to as "play calling," which enables the human operator to complete complex tasks through supervisory control and collaboration with intelligent automated agents. The following sections will therefore cover an overview of the concept of directability, an overview of the capabilities for action within this simulated air assault mission (e.g., aircraft payloads and capabilities), an in-depth discussion detailing TECUMSA's play calling interface, and finally a discussion of directability relative to the participant narratives from Chapter 4, where the participant narratives will be revisited through the lens of bridging Norman's (1986) gulf of execution (i.e., directability).

Designing Directability

> *"Directability means one's ability to direct the behavior of others and complementarily be directed by others. Directability includes explicit commands such as task allocation and role assignment as well as subtler influences, such as providing guidance or suggestions or even providing salient information that is anticipated to alter behavior, such as a warning."* p.52 (Johnson et al., 2014)

Simply put, directability is the user's authority, or ability to utilize affordances (Hutchins, Hollan, Norman, 1985; McDermott et al., 2018; Johnson et al., 2014). For example, a cell phone user may be able to direct their phone's consumption of power by changing battery settings to allow the phone to automatically put applications running in the background to sleep. However, if this customization setting is not available on their particular cellphone, the user does not have the authority/ability to direct that affordance (i.e., battery consumption rate), provided there are no other options in their battery settings of course.

User interface design acts as the medium for users to interact with the real-world affordances. This means that while affordances are the possibilities, directability (through the interface) is how you choose to navigate (or use) the possibilities. So, although the AMC is not able to change an aircraft's max airspeed (an affordance of the aircraft), they are able to choose or direct whether or not to fly the aircraft at that max speed. Ideally there is adequate transparency, or observability into the consequences of choosing to fly at max airspeed (e.g., quicker arrival to destination, but increased fuel consumption and thus decreased range of flight).

As a point of further clarification, *users are not creating affordances*, as it is the capabilities of aircraft for example that determine their affordances (e.g., max airspeed, hover ability, ability to shoot). Rather, it is *the AMC's ability to take advantage of those affordances that is the focus of this chapter* (e.g., change airspeed, select an armed rotorcraft aircraft for the task). Of course, the actions chosen may, in turn, shape the solution space – open up new opportunities and threats – but the gulf of execution is bridged by the directability of actions.

Affordances/Action Capabilities in Round 2

Table 4 (a duplicate of Table 3) contains a breakdown of capabilities for each airframe available to the AMC during the Round 2 Evaluation. A few of the most influential affordances of the aircraft were the *speeds* and *lethality* of each of the aircraft available to the AMC, which is consistent with the seven core competencies listed in the Army Aviation Field Manual No. 3-04 (2020, April). To help illustrate, Figure 29 contains a two-dimensional graph showing the relative speed-lethality plot for each aircraft available to the AMC in the Round 2 Evaluation.

Table 4. (Duplicate of Table 3) Aircraft Affordances in Vignette 4, Round 2 Evaluation.

ROUND 2, VIGNETTE 4

	Shadow (PS18)	Fire-X (DT39)	Gray Eagle (AE1)	Ownship (DN38)	ALEs (SR15 – SR19)
Camera Sensor	Equipped	Equipped	Equipped	Equipped	Equipped
Max Airspeed (knots)	70 kts	115 kts	134 kts	250 kts	80 kts
Radar Warning Receiver (aids threat detection)	N/A	Equipped	Equipped	Equipped	N/A
Missiles	–	4	2	4	–
Rockets	–	7	–	7	–
ALEs Onboard	–	–	2	2	–

*Note: All values are notional and for simulation purposes only, they are not reflective of any actual fielded weapon systems.

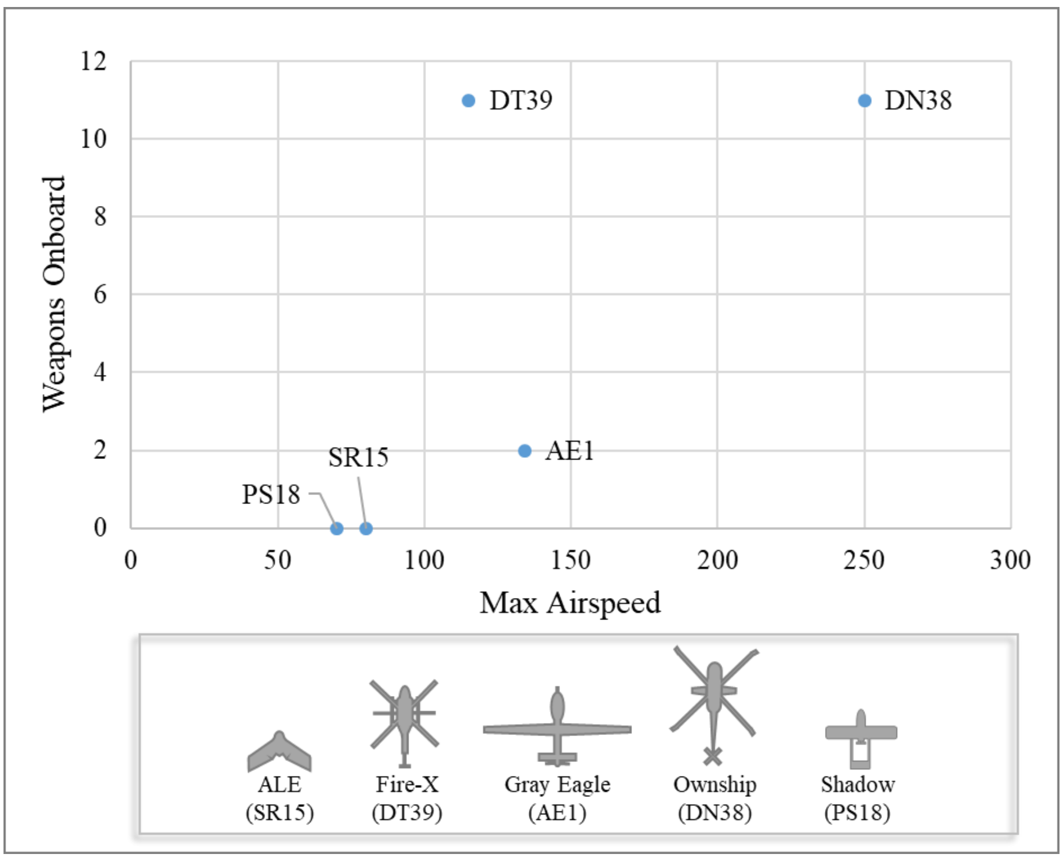


Figure 29. Speed-Lethality Difference (Max Airspeed-Weapons Onboard respectively) for each Aircraft in Vignette 4, Round 2.

Figure 29 is important because *operators must interact with "worldly affordances," not the limitations of the user interface*. Charles Sanders Peirce's Sign Theory, or Semiotics Theory, details the structure of thought and meaning processing as a triadic relationship between a *sign* (e.g., drawing of a cat), an *object* (e.g., the actual cat), and an *interpretant* (e.g., observer's interpretation of the drawing) (Flach, 2017; Yakin & Totu, 2014). This triadic relationship captures the importance of interpretation, which can be especially significant as we scale up to more complex problems.

Consider a sick patient seeking treatment from their physician. In this case, the signs are the patient's symptoms, family history, physical appearance, and so forth. The *meaning* of these signs is the actual illness/disease the patient is suffering from, and the *interpretant* is the physician's interpretation of the signs (i.e., the physician's diagnosis). Although ideally the *true meaning* of the signs and the *interpreted meaning* of the signs will align (i.e., the physician makes a correct diagnosis), this will not always be the case. We cannot fail to appreciate that a user's interpretation may not always mirror the true meaning.

It is therefore important that HMI designers *ground representations* in the practical realities of the physical ecology, and refine designs to ensure that the user's interpretation of the information aligns with the true state of the physical world the interface is intended to represent. For example, when the AMC encounters the following scenario seen in Figure 30 ideally their interpretation of the "fastest" aircraft aligns with the actual physics of these two objects traveling from different distances at slightly different speeds. Figure 30 shows a case where the advantage of using one aircraft over the other was more subtle ─ "Shadow" arrives 10 seconds sooner ─ since both aircraft have similar maximum airspeeds as well as proximities to the objective.

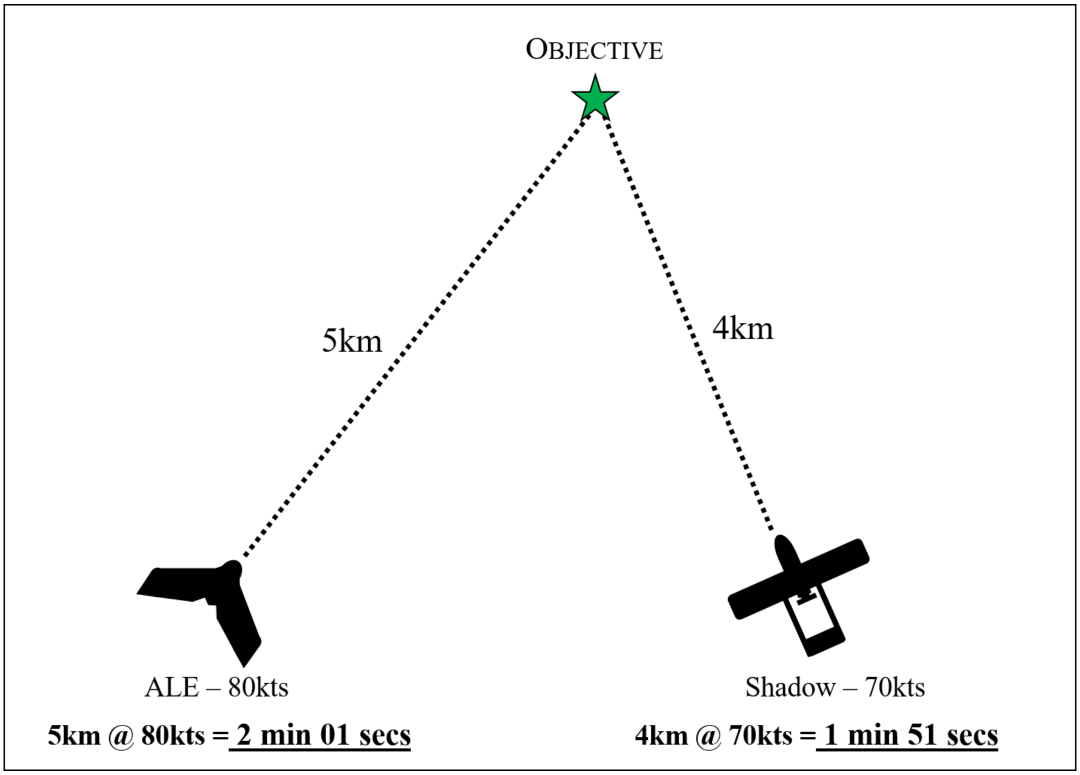


Figure 30. Illustration of Subtle Advantage between Two Aircraft that have Relatively Similar Maximum Airspeeds and Distances from the Objective.

So how does one approach the daunting task of determining what essential elements of information should be included in the user interface and/or the models used by autonomous systems? Of relevance to that question is Herbert Simon's (1996) parable of the ant on the beach, which postulates that *to understand the complex behavior of an ant travelling along the beach in search of food, one must first reflect upon the complexities of the environment in which the ant navigates.* There is a field of possibilities between the ant's "current state" and the "target state," or in Gibson and Crooks' (1938) terms a *field of safe travel*. To define the ant's field of possibilities or safe travel, a helpful start is first determining the constraints along the beach such as logs, boulders, and the shoreline. In other words, defining the constraints of executing an air assault operation provides a stable starting place for the work analysis, as well as the autonomous system development. For example, constraints affecting the flight characteristics of each aircraft (e.g., min/max/cruise airspeed, min/max altitude, safe/max bank angle) determine meaningful restrictions on feasible courses of action to the operator. *Simply put, when problem-solving is complex and there are multiple paths to the goal, it might be better to represent the constraints rather than the paths.*

Directability through Play Calling

In the real-world, a Commander's Intent is issued to subordinates, which is a broad description of what a successful mission would look like, with information such as the purpose of the operation (e.g., extend friendly forces further into hostile territory), and high-level broad guidance (e.g., time constraints for when the mission needs to be completed, the extent of resource availability). The reason for noting Commander's Intent is that it serves a parallel purpose as the technique of play calling.

In this case, the human operator serves as the "commander," and their ability to call plays is somewhat analogous to issuing a Commander's Intent statement to an autonomous subordinate. The play the user calls says *what should be done,* while the subordinate autonomous systems of TECUMSA have the authority to decide *how it should be done*. Distributed control is a common architecture in more domains than the military, including project managers for a company, or physicians working with a team of nurses. Quite frequently, not every task/sub-task can be completed by one individual, so giving distributed decision-making authority across the team enables far greater adaptability than a hierarchical decision-making process that requires negotiating across a chain of command before any formal action can be taken.

> *Directability is supported when humans are able to easily direct and redirect an automated partner's resources, activities, and priorities. Humans will have expertise or insights that the automated system will not. Humans are ultimately accountable for system performance, which means they must be able to stop processes, change course, toggle between levels of autonomy, or override and manually control automation when necessary.* p.31 - McDermott et al. (2018)

Orchestrating several unmanned and manned aircraft in a contested battlespace is easily an overwhelming task for a single human to be controlling. Consider the number of user inputs that would be required to manually generate a flight route for even a single aircraft. A significant amount of time would be needed to calculate the shortest path between two points, as well as determine how to best deconflict with terrain (e.g., flying around vs over a mountainous ridgeline), all while computing the required standoff that would provide ideal camera sensor coverage and/or achieve weapons in-range while avoiding NFAs. At each waypoint along the route the user would need to compute the acceptable descent/ascent rate for that aircraft based on its unique flight characteristics, the crosswinds, the air pressure, the weather, and so forth. In fact, entire careers are dedicated to determining how to best generate a flight route for a UAV (see Stecz & Gromada, 2020 for additional details on UAV mission planning for reconnaissance applications). *At this level of direct involvement, the user has complete control over the system, but this is not efficient*.

On the other end of the continuum of control, the flight routes could be generated autonomously by intelligent algorithms. This would involve circumventing the limitations of the human operator by removing them from the decision-making loop so that the autonomy could be in charge of generating a flight route based on known aircraft flight capabilities, sensor limitations, and terrain characteristics. Although the computer processing power and algorithms are available to quickly generate a fast and usable route for each aircraft to fly, a different challenge presents itself. That is, the autonomy would struggle to generate an *appropriate* route without being given a purpose for the route in the first place.

Consider that in a dynamic, contested battlespace the goal for flying a route is not always the same. Sometimes, a route is flown so an aircraft can quickly arrive at its destination to complete an objective, other times the goal is to travel more stealthily to arrive at the destination without being detected, and in other cases the goal is to collect intelligence as the route is flown. The best flight route to complete a given task, and the best aircraft to complete the task, both depend entirely on the goal of the AMC and the priorities and constraints of the task at hand. *The play calling approach is therefore designed to let the AMC specify the command intent, but leaves the autonomous system to handle many of the specific decisions about how to achieve that intent.*

The TECUMSA user is presented with a bank of high-level tasking options, also known as “plays,” to quickly communicate to the autonomy a higher-order objective (e.g., "I want to look somewhere, "I want to follow something", “I want to kinetically engage something”). With that information, the autonomy is able to make a recommendation to the human for what aircraft it thinks would be best to achieve that goal, and plan a route for the given aircraft to fly. For example, by calling a *Parallel Search* play the operator is directing the autonomy to conduct reconnaissance on a larger area, at which point the autonomy can then recommend an aircraft such as the Grey Eagle UAV, perhaps because it is currently available and has a high range sensor that can quickly scan a large area from further away.

Play calling is an elegant solution that streamlines the user inputs needed to enable quick, high-level coordination of goals, priorities, and activities to an autonomous teammate. Play calling also facilitates increased adaptability, as the human is able to dynamically jump in to direct and redirect the courses of action as needed when unexpected situations occur (e.g., the sudden ambush of Ground Troops under fire).

The use of a play calling interface approach allows the user to quickly say “what” needs to be done, but also provides *optional* modifications so the operator can further specify “how” it should be done. More specifically, after a play is called a Play Workbook opens for the user to fine-tune and adjust the play parameters as needed. The user then has the option to modify the play as little or as much as they want depending on the situational needs (see the *TECUMSA: Play Calling Process* section for more details). Notably, the Play Workbook chunks information into meaningful units to promote user interaction at an appropriate level to guide the autonomy (i.e., not too much detail, nor too little). This includes value parameters to adjust aircraft selection (e.g., leaning towards a more heavily armed aircraft, requiring an aircraft that can hover), as well as flight parameters such as the location/type of loiter when it reaches its destination.

Timeline Narratives

As previously mentioned, the following section will cover the timeline narration of events (duplicated from Chapter 4) and will be followed by a discussion of directability relevant to that passage of the narrative. Italicized font and text indentation will be used to separate the original narratives (pulled from Chapter 4– Vignette 4 Timeline Narratives) from the additional commentary that focuses on directability. More specifically, text that is indented and italicized indicates the participant's original detailed narration pulled from Chapter 4, while the text closer to the margins and un-italicized is the new expanded upon details of directability.

Participant 7's Detailed Narration - Directability

> *Participant 7 begins their course of action by having their "Fire-X" UAV and Ownship reconning Route Jarus, and their "Shadow" and "Gray Eagle" UAVs reconning NAI 26; these aircraft remain on these tasks for the first quarter of the 60-minute vignette. At 04:00 minutes into the vignette, Participant 7 receives a chat message with information from the first pop-up event reporting heavy enemy activity near the Ground Troops to the East of the dam. At 04:04 the participant positively identifies a group of hostile enemy armored personnel carriers (APCs) in NAI 26, which they choose to kinetically engage twice over the next seven minutes. Although the participant's attention was primarily focused on engaging the hostile APCs, mid-way through this engagement the participant briefly turns their attention over to read their chat message (at 06:25), 2 minutes and 25 seconds after receiving it.*

In the *TECUMSA: Human-Autonomy Teaming During Engagements* section, a description of this typical engagement process is provided. Key behaviors or elements of directability occur at each step of the process, from choosing what play (recon-based versus engagement-based), to choosing how and when to kinetically engage the threat. At a later point in Participant 7's narrative (specifically around minute 26:20), additional details about human-autonomy coordination and task hand-offs will be discussed.

> *At 14:00 minutes into the vignette, the Ground Troops report via chat that an enemy patrol is approaching their position (the second pop-up event), which the participant reads 2 minutes and 6 seconds later. The previous two chat messages (at 04:00 and 14:00 minutes) were foreshadowing of the next chat message at 16:00, which is a critical turning*

point in the scripted mission priorities. At 16:00, the Ground Troops report via chat that they are now under attack (the official start of the third pop-up event, which is the most mission impactful) and request immediate armed assistance from the participant. Simultaneous to this incoming chat message, two audio/visual alerts of gunfire detection occur, along with a large orange icon in the map indicating the precise location of a gunfight that the Ground Troops are now in with an enemy patrol.

Up until the time of the gunfire detection at 16:00, there have been 6 missiles launched at one or more of the participant's team of aircraft. Note the participant is not told which aircraft is getting shot at, only the grid location the enemy missile launched from. In the TECUMSA system, enemy missile launches were accompanied by both a salient auditory warning as well as a large icon in the map indicating the location of the enemy missile launch. Despite the saliency and time pressure of these missile launches, the audio/visual alert of the gunfire was prominent enough to grab Participant 7's attention. Shortly after noticing the gunfire, Participant 7 quickly tasks Fire-X to provide support to the Ground Troops.

The "quick tasking" of Fire-X to provide armed support to the Ground Troops is a critical gulf in directability that will be discussed in much further detail in the upcoming paragraphs. It is important to note that Participant 7 used the typical play calling process to direct Fire-X to the location of the Ground Troops. Unfortunately, the time the participant gained by quickly tasking the aircraft they then lost due to the suboptimal aircraft selection for that task. The famous quote by Benjamin Franklin comes to mind here, that "an ounce of prevention is worth a pound of cure."

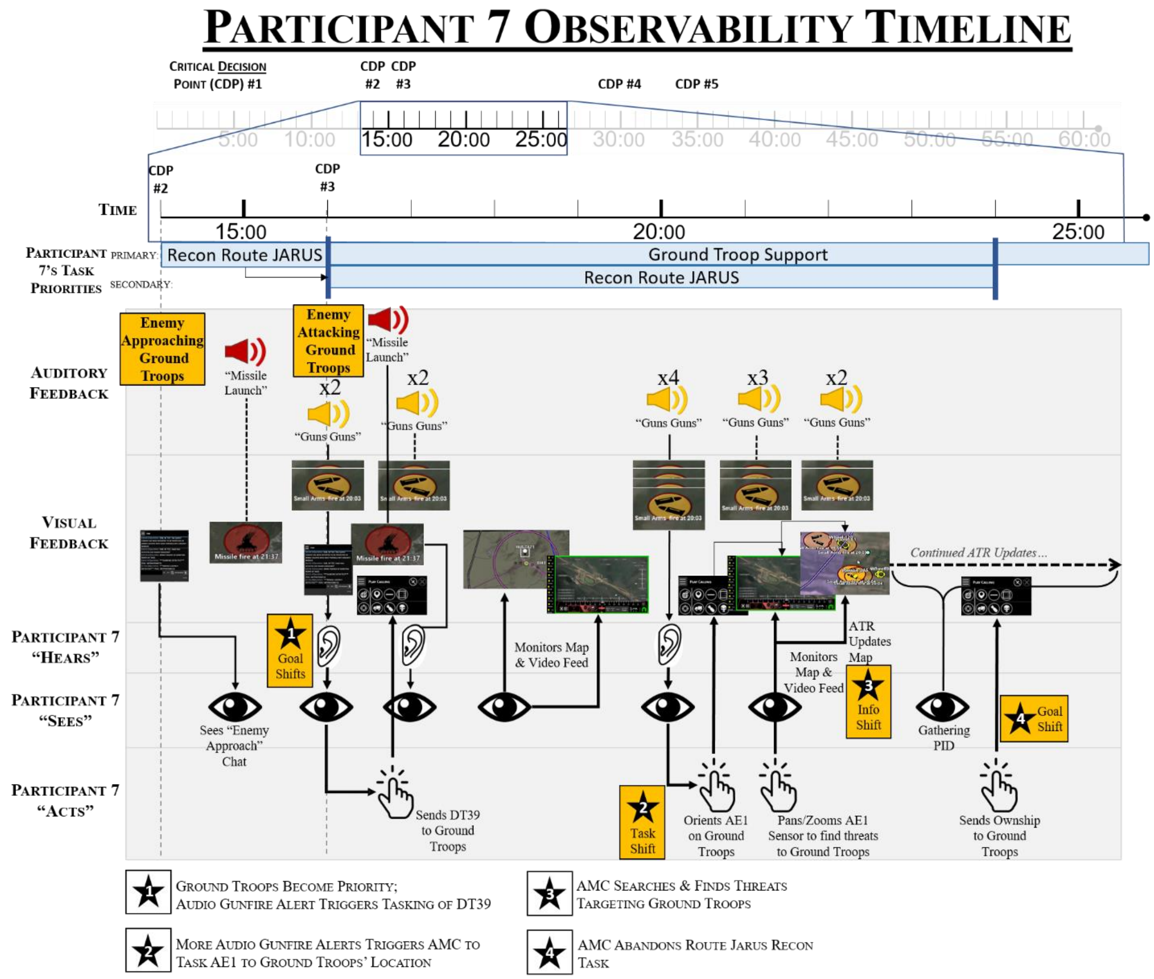


Figure 31. (Duplicate of Figure 22) Deconstructed Timeline for Participant 7 from Minute 14:00 to 26:00 in Vignette 4, Round 2 Evaluation.

> *At 20:41 the participant realizes how long it would take for Fire-X to reach the Ground Troops, and decides to also task the Gray Eagle to provide armed support to the Ground Troops under attack. Because the Gray Eagle was already in range of the Ground Troops, the participant was able to immediately start searching for entities in the location of the gunfire in hopes of finding the threats attacking the Ground Troops. After searching the map and video feeds for entities near the gunfire for 3 minutes and 27 seconds, the participant finds a few wheeled vehicles in the location of the gunfire but has not yet been able to positively identify them as friendly or enemy.*

In the case of assisting the Ground Troops, there were two critical constraints that mattered for determining an acceptable solution: Time taken to reach the Ground Troops, and the need for an armed aircraft. As previously discussed in Chapter 5 at this same point in Participant 7's vignette narrative, there lacked sufficient transparency during the play calling process into how long a particular aircraft was expected to take to reach its destination. However, a noteworthy directability misalignment is also at the root of this less than ideal tasking of Fire-X to help the Ground Troops. During the post vignette questionnaires, it appeared many users had a different weighting schema for aircraft desirability than the autonomous allocator agent. More specifically, the allocator agent strongly preferred to avoid having to pull aircraft off of an ongoing task, and thus prioritized *availability* of aircraft as more important than the estimated flight time of an aircraft.

While this prioritization schema generally worked (i.e., first considering the aircraft that were un-tasked), there are exceptions to this rule. One such exception was revealed when the Ground Troops came under hostile fire. In that case, many of the participants felt that the most important factor in deciding the best aircraft was merely finding the armed aircraft with the shortest estimated flight time. More specifically in this case, participants were typically willing to suspend *any* ongoing task to maximally reduce the time needed to arrive and support the Ground Troops. Participant 7's late realization at 20:41 that Fire-X was not going to arrive in an acceptable amount of time to the Ground Troops meant that *the user was unable to effectively coordinate their goals with the autonomous allocator agent to guide its recommended aircraft tasking solutions to assist the Ground Troops.*

This tasking of Fire-X is key to the discussion of bridging directability. The AMC made a choice that was incompatible with their intention. Furthermore, the play fulfillment algorithm recommended a solution that was incompatible with the AMC's intention. It is important to note that the Play Workbook tile *did provide the opportunity to consider tasked vehicles*, the AMC just did not utilize this capability, which was offered in the Quick Swap tile (see the *Play Workbook: Quick Swap Tile* section). Based on lessons learned in the Round 1 Evaluations, designers added the ability in the Round 2 Quick Swap tile for users to view the allocation algorithm's recommendation for aircraft tasking *independent* of whether or not aircraft were already busy on an ongoing play. However, this is a great example of the importance of a rich coupling between observability and directability, because the user gaining transparency into the full solution space (through the Quick Swap tile) would have likely been more frequent if the information was less effortful to access (e.g., fewer clicks to access a feature should increase the likelihood of its use).

> *Around 24:00 the participant notices that Fire-X is still several minutes away from the Ground Troops. Since Gray Eagle has only 1 missile remaining in its weapons payload, the participant shifts their attention to getting additional armed support over to assist the Ground Troops. At 24:08 the participant pulls the Ownship off the task of reconning Route Jarus and instead directs it to head over to support the Ground Troops, which becomes their third of four available aircraft (excluding the ALEs, which are not launched) tasked to assist the Ground Troops up to this point in the vignette. After tasking the Ownship, the participant returns to looking for the enemies engaging the Ground Troops through Gray Eagle's camera sensor.*

At this point in Participant 7's vignette, they had abandoned three ongoing reconnaissance tasks so they could send those three aircraft over to assist the Ground Troops that were still under hostile fire. Each of these re-tasking decisions subsequently left three priority reconnaissance tasks unfinished. However, at 24:00 minutes into the vignette, Participant 7 still had available resources. Specifically, they still had four available ALEs each with a camera sensor, which had yet to be launched and were thus sitting unused onboard the Grey Eagle and Ownship aircraft (two ALEs on each). This re-tasking strategy highlights that there were gaps in sensor coverage for the reconnaissance tasks, and this coverage could have been significantly increased by using the four available ALEs.

In Chapter 5 support for observability was proposed so that future autonomy development would include user interface improvements so that *unlaunched* ALEs could still be considered by the allocator agent as viable solutions to use for play tasking. Since perception is tightly coupled with action, this improvement to observability is a clear segue to expand the solution space for the user to navigate through (a feature of directability). More specifically, including additional aircraft with varying capabilities as viable options to consider for play calling would ultimately enhance the user's ability to direct system performance. In practical terms, this adjustment would have meant that rather than pulling aircraft off of tasks, the participant would have been prompted to launch an ALE instead.

> *At 26:20 the participant is able to positively identify the three hostile enemy technical vehicles attacking the Ground Troops, and then reclassifies the threats as "hostile" in the TECUMSA system. At 26:30 the participant calls an Engage Play using the Gray Eagle, and at 26:55 the participant launches missiles off of Gray Eagle at the hostile threats. At 27:20 the participant confirms the hostile technical vehicles are destroyed and has therefore successfully supported the unanticipated pop-up event of ambushed Ground Troops.*

One key aspect of directability is supporting role allocation between the human and the autonomous agent. However, most of the responsibilities of the human and the autonomous agent in TECUMSA are static. This means that although there are a significant number of tasks that the human "hands-off" to the autonomy (and vice versa), each hand-off is always triggered by the same event (e.g., once the autonomy positions the aircraft in range for weapons release, the human must choose which weapon to confirm for weapons release at the threat). One of the most notable coordinated tasks between the human and the autonomy's activities occurred during engagements. Table 5 is the typical workflow from start-to-end of an engagement play, as well as the task allocation between the human and autonomy.

Table 5. Human-Autonomy Task Allocation.

Task Description	Human's Task	Autonomy's Task
1. Identify entity as a threat	✕	
2. Mark threat as "hostile"	✕	
3. Call Engage play	✕	
4. Recommend aircraft for play		✕
5. Plan Flight Route		✕
6. (Optional) Adjust play details (e.g., change aircraft selection)	✕	
7. Confirm engagement plan	✕	
8. Control aircraft's heading, altitude, & route following		✕
9. Confirm threat position & identification before weapon launch	✕	
10. Ensure aircraft is in range for weapons release		✕
11. Choose weapon / confirm actual weapon's release	✕	
12. Maintain eyes-on target with camera sensor		✕
13. Conduct BDA after weapons release	✕	
14. Determine whether to fire another weapon or end play	✕	
15. (Optional) Mark threat as "destroyed"	✕	

This shared tasking, or back-and-forth of authority, highlights the collaborative process between the human and the autonomy. Keep in mind that the human has the ultimate authority at any point to cancel the play altogether, as well as pause the play and re-direct the goals as needed. Furthermore, there are steps where the autonomy asks for the human's final approval, for example when moving from Step 6 to Step 7 the user has the option to adjust play details before confirming play execution. This role allocation scheme allows the human to retain ultimate command authority (i.e., directability), but remain "in-the-loop" during play execution, which is particularly important during higher consequence plays such as kinetic engagements.

> *After neutralizing these threats, the participant returns to trying to precisely locate the Ground Troops location and communicates with the TOC (played by an experimenter) to get information on the status of the Ground Troops. The participant specifically asks the TOC over radio (via headsets) communication, "Are they [Ground Troops] alive?" Behind the scenes the TOC stalls to answer the participant's question, because in 55 seconds the fourth critical chat message is scheduled to be sent to the participant (at 30:00), requesting a medical evacuation (MEDEVAC) for the Ground Troops. In the meantime, as the participant waits for status information on the Ground Troops, they work to find threats from the surrounding areas of the Ground Troops' approximate location.*
>
> *At 30:00 the fourth pop-up event initiates, where a chat message is sent to the participant requesting MEDEVAC support for the Ground Troops. At this point, the Ownship is still enroute to the Ground Troops. At 33:41 the participant checks their chat window and sees the Ground Troops' request for a MEDEVAC. By the time the Ownship reaches the Ground Troops location, at 34:01, four missiles have been launched at one or more of the participant's aircraft. It is at this arrival time at 34:01 that the experimenters notice the Ownship aircraft hovering just above the ground and barely moving forward. This was because, upon arrival to the Ground Troops, Participant 7 had the pilot position the Ownship to land and actually carry out the MEDEVAC. It is worth noting that no other participant did this, and behind the scenes experimenters were extremely*

> *impressed by the participant's dedication to help the injured Ground Troops.*
>
> *After the participant offers to provide the Ground Troop's MEDEVAC, 59 seconds go by. The fifth and final pop-up event, which is a scheduled chat message, is then sent at 35:00 stating that a MEDEVAC aircraft team is 25 minutes from the current Ground Troop's location and will fly ingress via Route Secura and land in LZ Soontir. Since Participant 7 has their Ownship positioned to actually carry out the MEDEVAC, the TOC adapts the script by further clarifying to the participant over radio communications to "Charlie Mike" (or Continue Mission) because the Ground Troops need specialized medical equipment.*
>
> *At 35:00 the participant turns their efforts towards clearing the surrounding area of threats. After 4 minutes and 36 seconds of searching, the participant positively identifies an anti-aircraft missile system (hereon referred to as a "hostile missile system"), which has been launching several missiles at the participant's aircraft throughout the mission. At 39:58 the participant sends the first LRPF at this threat (i.e., calling for a missile launch from a remote ground location), and after not observing the desired kinetic effects, the participant decides at 48:20 to send a second and final LRPF at the hostile missile system. Shortly after the second LRPF, the participant confirms the hostile missile system is destroyed through a nearby aircraft video feed and marks it as destroyed in the TECUMSA map.*

In actual military operations there are a variety of factors to consider in selecting a particular weapon system (e.g., target type, desired effects, quantity of weapon systems, time taken for the weapon to reach the target). However, the most noteworthy aspect for directability is the user's ability to choose a particular weapon system. The TECUMSA system for Round 2 was designed to allow the user to adapt their use of different weapon systems depending on the situational conditions. Because kinetic engagements were such a critical step in this synthetic task environment (as well as in real-world operations), designers made sure the user was able to precisely direct their chosen weaponeering solutions (i.e., missile, rocket, LRPF), as discussed earlier in Figure 14.

> *At 49:00 into the vignette, 14 minutes after being notified of the new route and landing zone needing to be reconned (i.e., Route Secura and LZ Soontir) for the Ground Troop MEDEVAC, the participant still has three of the four aircraft loitering nearby to the Ground Troops. In other words, the participant has not tasked any of the aircraft to specifically search along Route Secura or LZ Soontir. At 56:16 the participant finds one of the APCs in NAI 26 that had not previously been impacted by the first engagement. At 56:19 the participant sends an LRPF request, and at 56:40 confirms the APC threat has been destroyed.*
>
> *At 60:00, the end of the vignette, the TOC asks the participant whether or not they advise the MEDEVAC operation to continue. Since the participant had neutralized the known hostile missile system and did not detect or know of any other threats that would impact the incoming MEDEVAC team, they gave the TOC an affirmative "Go" call to continue with the MEDEVAC. After consulting with SMEs after the completion of this run, both SMEs observing the vignette agreed that this was a good call made by the participant.*

Participant 7 Directability Summary. *Directability* in interface design is focused on the authority to specify an intention/command and to initiate action. This includes updating autonomy to the dynamic flow of resources, activities, and priorities, as well as making direct inputs such as launching an ALE or confirming weapons launch at a threat. In reviewing directability in Participant 7's Vignette 4 run, it is most helpful to think of directability as falling into one of the four quadrants seen in Table 6: (I) Had directability authority and used it to influence actions, (II) Did not have the authority, but needed the ability to influence actions, (III) Had the authority, but did not use it to influence actions, and (IV) Did not have the authority and did not need to use it to influence actions.

Table 6. Directability Matrix.

Used / Acted		Authority: Yes	Authority: No
	Yes	Quadrant I Legitimate Choice	Quadrant II Intent Violation or Workaround
	No	Quadrant III Missed Opportunity	Quadrant IV Legitimate Discretion

Two of the quadrants in Table 6 are of particular interest to the topic of supporting directability in interface design, specifically Quadrants (II) and (III) as seen highlighted in blue. The first gulf in directability emerged when *the participant had the authority/ability to direct actions but did not use it* — unfortunately to the detriment of their performance. As denoted in Quadrant III, this type of directability gulf is considered a *missed opportunity* on behalf of the participant. One key example of this was Participant 7's suboptimal tasking of Fire-X to support the ambushed Ground Troops; they *could* have tasked Grey Eagle, but disuse of the Quick Swap tile meant they did not choose the fastest aircraft because it was already busy. Another important example of this occurred when Participant 7 neglected to launch any of their available ALEs, despite having the authority/ability to do so, which greatly hindered their ability to complete all reconnaissance tasks.

The second noteworthy gulf in directability occurred when *the participant did not have adequate authority/ability to direct actions but needed the ability*. As denoted in Quadrant II, this type of directability gulf resulted in an *intent violation or workaround behaviors* on behalf of the participant. One very noteworthy example of this gap in coordination/directability between the participant and the autonomy was revealed in the differing decision-making criteria used by the autonomy versus the human when determining the "best" aircraft to allocate to a given play task.

Depending on the circumstances, the human operator would sometimes value *arrival time* as the most important criteria, whereas the autonomous allocator agent would instead value aircraft *availability* as the most important. More specifically, the allocator agent strongly preferred to use an already available/untasked aircraft, rather than have to pull an aircraft off an ongoing play (provided the available aircraft had the minimum required capabilities). This misalignment in selection criteria meant that there were times when the "best" aircraft for the play was not agreed upon by the human and the autonomy, which was observed in Participant 7's Vignette 4 run when the Ground Troops required immediate armed support and the Fire-X aircraft did not arrive as soon as desired.

An additional directability gulf was also at the root of this periodic misalignment between humans and autonomy when choosing the "best" aircraft for a play. Specifically, it became clear that additional directability was needed for the user to coordinate the *priority* of a given task. Since there was no way for the participant to communicate with the autonomous systems regarding the intended criticality of each task, it was not possible for the autonomy to intelligently choose which plays to suspend so that *higher* *priority* tasks could be completed. To improve directability in future interface designs, it would likely help to allow the user the ability to mark a task as either *primary*, *secondary*, or *tertiary* criticality levels.

Discussion – Participant 7 Directability. A noteworthy element of directability involves participant use of the Quick Swap tile. As seen in Figure 13, TECUMSA provided a Quick Swap tile that allowed the user to review different options to complete a particular play (e.g., what are the advantages of using aircraft X over aircraft Y). To support the user's ability to explore the entire solution space, the tile was modified for Round 2 to include a selectable option to "Show Busy Vehicles," so that both available and busy aircraft could be considered by the allocator agent for the play at hand. This HMI change was used a considerable number of times during the Round 2 Evaluation. The tile itself was opened 240 times, the "Show Busy Aircraft" option was selected a total of 57 times, and an alternative play was chosen 26 times.

While it is validating to see a particular design feature being used, there is a more valuable data point to extract from this finding. Using the Quick Swap tile can be a way to further explore the solution space, but when an alternative play is chosen, that is actually a strong indication that there was a tuning misalignment between how the operators wanted to complete plays and how the allocator agent thought they should complete plays. As previously described, the leading reason at the root of this misalignment is that the human and the autonomy did not always share the same decision-making criteria when selecting aircraft for plays (e.g., user prioritized the fastest arrival time, whereas autonomy prioritized currently available/untasked aircraft).

A couple of additional explanations are possible for users regular use of the Quick Swap tile to select an alternative play execution strategy to the one originally recommended by the autonomy (e.g., using the Grey Eagle for the engagement play on "Threat 444" rather than Fire-X). The main reason is likely because the human already had the desired aircraft in mind. If that were the case, it is possibly quicker (fewer mouse clicks, less cognitive effort) to search through the Quick Swap play options for the desired aircraft allocation, rather than choosing multiple optimization parameters in the play workbook to get the desired aircraft to be recommended for the play. In either case, (a) a misalignment in deciding the "best" aircraft for a play or (b) the Quick Swap tile was a quicker way for users to find their desired aircraft, it is an opportunity to improve directability in the interface and goal coordination between the human and autonomy.

Participant 2's Detailed Narration - Directability

> *In the first five minutes of Vignette 4, Participant 2 has their "Fire-X" UAV tasked to recon NAI 12, their "Gray Eagle" UAV reconning TAI 63, and their "Shadow" UAV and Ownship tasked to do a joint recon of Route Jarus. At 04:00 minutes into the vignette, Participant 2 receives a chat message pertaining to the first pop-up event reporting heavy enemy activity near the Ground Troops to the East of the dam, but the participant never read this chat message. At 06:29, the participant pulls the Shadow off of the joint recon of Route Jarus after noticing that it will take a significant amount of time for the Shadow just to travel across the battlespace and start reconning Route Jarus. The participant then instead tasks the Shadow to recon NAI 11, which was a much closer area to the starting location of the Shadow.*

Within the first five minutes of the vignette, Participant 2 was able to efficiently task their four primary aircraft. Experimenters behind the scenes actually noted how comfortable this participant seemed navigating the TECUMSA system; while there was a full day of training on how to use TECUMSA, it is still a complex interface that not all participants naturally picked up. It is also worth exploring the keen awareness Participant 2 showed when they pulled the Shadow off of the joint recon of Route Jarus.

As previously commented on, there is information that could be presented during the play calling process in future designs of TECUMSA that would increase the transparency of projected aircraft flight times *before* the participant commits to executing the play. However, it is worth noting that this Participant seemed to already be aware of the general action possibilities for each of their aircraft (e.g., clearer mental model for converting airspeed in knots to flight times), as they checked in on the progress of the Shadow after 2 minutes and 29 seconds, at which point realizing they needed to re-direct that aircraft's tasking assignment. This highlights the importance of knowing who you are designing the interface for, since there is a significant interplay between user experience/knowledge and their learned ability to command and direct the flow of resources in this complex battlespace.

> *At 06:42 the participant launches the first ALE off of their Ownship, and then at 06:51 launches the second ALE off of the Ownship. With the two available ALEs on the Ownship launched, the participant then tasks both aircraft to inspect the location of a templated hostile missile system. Once the two ALEs are tasked, the participant turns their focus to finding hostile threats in the battlespace.*

Launching the ALEs was a very quick process. First, the operator would click a button in the Vehicle Dashboard tile to "Launch ALE" (see Figure 32), and then press a second button to confirm launch (see Figure 33). Once confirmed, the ALEs would automatically launch and become an additional asset the participant could utilize to accomplish mission tasks.

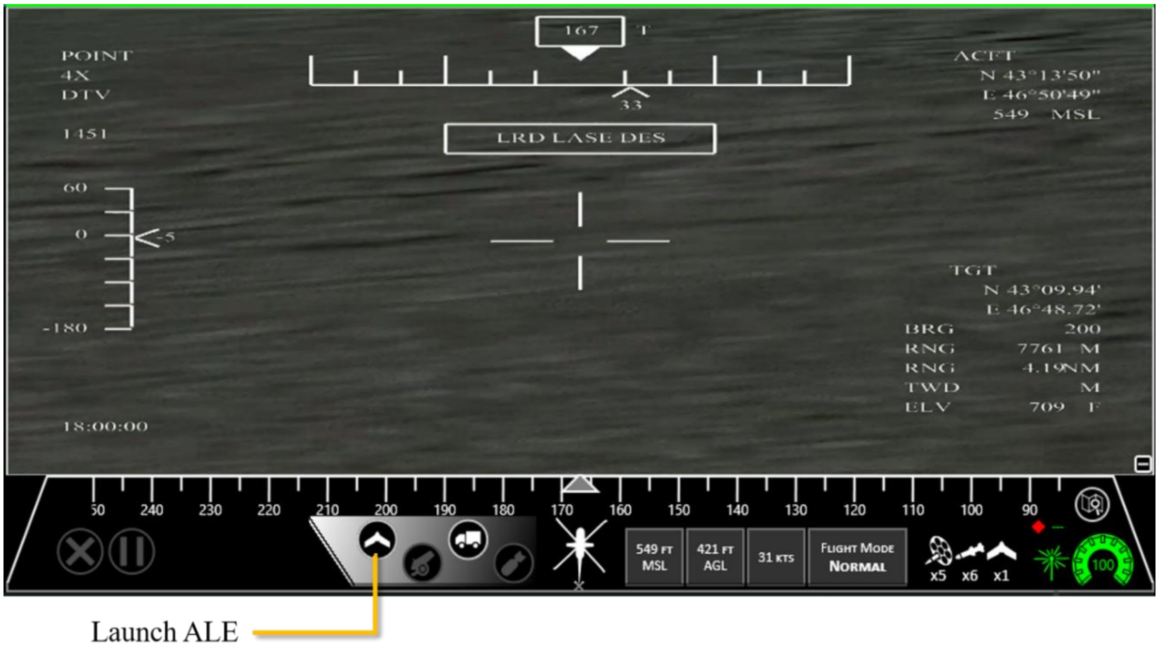


Figure 32. Vehicle Dashboard with "Launch ALE" Button Callout.

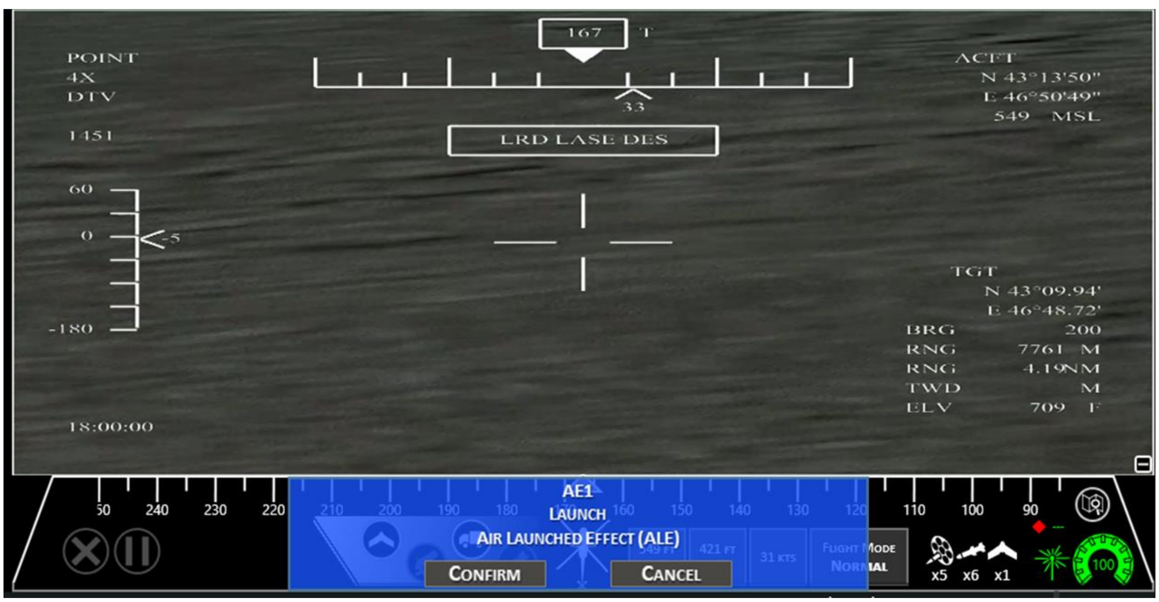


Figure 33. Vehicle Dashboard with the Confirmation/Cancel Window After the "Launch ALE" Button is Clicked.

> *At 13:49 the participant identifies a wheeled vehicle near the vicinity of the templated hostile missile system. At 14:00 minutes into the vignette, the Ground Troops report via chat that an enemy patrol is*

> *approaching their position (the second pop-up event), which the participant never reads. After confirming no friendlies are in the nearby location to the templated hostile missile system, the participant calls an LRPF on a wheeled threat at that location at 15:06. At the exact same time (15:06), a gunfire detection audio/visual alert is generated. Five seconds later at 15:11 the third critical chat message is sent, reporting that the Ground Troops are now under attack and request immediate armed assistance from the participant (the official start of the third pop-up event, which is the most mission impactful).*

To strike the wheeled threat at the hostile enemy compound, the participant first had to reclassify the entity from "unknown" to "hostile." This was typically accomplished by right-clicking on objects in the map to spawn a radial menu (Figure 34). This is a particularly important step, because multiple system features are blocked until the participant verifies the entity is hostile. For example, an Engage Play cannot be called against an entity until it was marked as "hostile" by the participant. This reclassification constraint actually utilizes an important concept to directability: *necessary friction* (Rochlin, 1997; Rochlin, La Porte, & Roberts, 1998).

Although all human-computer interfaces are intended to increase user productivity, Rochlin (1997) cautions that a system that makes it easier to do the right thing may also make it easier to do the wrong thing. For example, an input error with just one decimal point off or the accidental addition of a zero can dangerously change the amount of medication a patient receives. Rochlin's work analyzing features of high-reliability or nearly zero error rate organizations, such as U.S. Navy aircraft carriers and air traffic control, is to have resilience built into the system.

More specifically, resilience can be seen in human-computer interface design where the goal is not to universally make all tasks easier. Some tasks, such as kinetically engaging enemy threats, should be met with *necessary friction* (e.g., requiring "hostile" classification, message prompts requiring confirmation of weapon release). Interface design techniques that employ confirmation prompts, error messages, warning messages, rejections of invalid inputs, and cross-validations (e.g., user-inputted “A” here, but then “C” there) can all be used to slow down the speed of user actions in an attempt to reduce errors and thus create necessary friction.

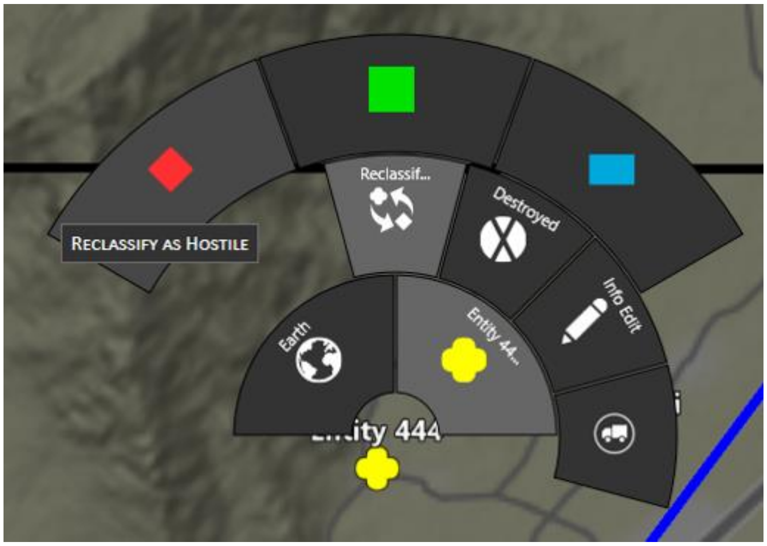


Figure 34. Radial Menu Used to Reclassify Entities as Hostile.

> *Although the participant does not immediately read the third chat message reporting the Ground Troops under attack, the audio/visual alert of gunfire at 15:06 was salient enough to draw their attention to the Ground Troops location on the map. Up until this point, the participant has not read any of the critical chat messages tied to the scripted pop-up events in Vignette 4. However, Participant 2 correctly decides that the gunfire is potentially coming from and/or directed towards the Ground Troops that were known to be roughly in the area of the gunfire, based on their pre-mission briefing.*
>
> *At 16:13 the participant has the Ownship head towards the gunfire, but at a safe distance to the South. After redirecting the Ownship and pulling it off of the Route Jarus reconnaissance, the participant checks their chat tile for additional information in regards to the gunfire detection. At 17:10 the participant reads that the Ground Troops are requesting immediate armed support from the participant. At 17:24 the participant*

sees multiple wheeled vehicle icons pop up in the map at the location of the gunfire. Participant 2 then spends the next 38 seconds confirming that the newly detected wheeled entities are indeed hostile and not the Ground Troops fighting back.

One noteworthy influence on user behavior came from the teaming between the participant and the “human autopilot” flying the Ownship aircraft. In both the Round 1 and Round 2 Evaluations, the participant could issue commands (e.g., tasking, flight maneuvers, changes in altitude or airspeed) to their Ownship pilot through two different methods: verbally over radio headset, or digitally through the TECUMSA interface (i.e., play calling to generate a route for the human Ownship pilot to follow).

In this case, Participant 2 chose to verbally communicate with the Ownship pilot (a human operator), to issue the command to fly South of Route Jarus towards the recent gunfire detection, with an additional note to “…use terrain masking on the ingress flight approach.” It appears that this verbal command to the Ownship pilot was used as a streamlined way of directing behaviors of the Ownship, since any digital flight commands that were issued were still dependent on the Ownship pilot following them. In other words, verbally directing the Ownship pilot was essentially “cutting out the middle man.” Because verbally issuing commands to the Ownship pilot was commonly observed across participants, there may be additional benefits to this modality of issuing commands worth exploring in the future (e.g., speech recognition capabilities).

From 18:02 to 20:43 the participant is directing the Ownship pilot to get in position to engage the hostile threats attacking the Ground Troops. From 20:43 to 23:04 the participant is confirming the location of the Ground Troops to make sure they are not in close proximity to the intended kinetic strike location. At 24:10 the participant calls an Engage Play using the Ownship, and at 24:30 the participant launches missiles off of their Ownship at the hostile threats. At 25:19 the participant confirms the hostile

> *technical vehicles are destroyed, and has therefore successfully supported the unanticipated pop-up event of ambushed Ground Troops.*

It is worth noting that the process of engaging threats is the same when using either the Ownship or a UAV. However, in this particular case Participant 2 was a bit more hands-on in the positioning of their aircraft prior to weapons release. This makes sense since there are a variety of flight tactics to reduce risk to the firing aircraft, such as hiding behind nearby terrain to avoid return fire from the enemy. These tactics are far more paramount when engaging with a manned aircraft as opposed to an unmanned aircraft, which is a notable directability requirement to consider in future designs of TECUMSA.

> *Following this engagement, at 25:53 a missile is launched at one of the participant's aircraft. At 26:25 the participant directs their attention to the location of the missile launch site. At 27:17 the participant identifies the missile launch site as hostile and at 27:27 sends a request for LRPF at the location of the enemy missile launch. From 25:53 to 27:49, four missiles in total are launched at the participant's aircraft. At 29:12 the participant learns that the Fire-X UAV aircraft has been shot down by an enemy missile (i.e., one of the enemy missile launches hit its target). At 30:00 the fourth pop-up event initiates, where a chat message is sent to the participant requesting MEDEVAC support for the Ground Troops.*
>
> *With Fire-X no longer available for tasking, the participant returns to their task of finding threats in the battlespace. From 31:58 to 37:50 the participant focuses on finding threats in TAI 63. At 37:50 the participant reorients their attention back to templated hostile missile system. At 38:51 they focus specifically on the icon in the map for Wheeled Entity 152, which is in proximity of the templated hostile missile system. However, in actuality the Wheeled Entity 152 has moved from that location, and the color faded icon in the map only indicates the last known location. The participant misinterprets what the icon in the map indicates and decides to pursue an engagement of the Wheeled Vehicle 152 due to its proximity to the templated threat (i.e., the hostile missile system). At 41:00 the participant calls an Engage Play on Wheeled Entity 152 using the Gray Eagle.*

At 42:47 the fifth and final pop-up event initiates, where a chat message is sent to the participant stating that the MEDEVAC team is 25 minutes from the current Ground Troop's location and will ingress via Route Secura and land in LZ Soontir. Eight seconds after receiving the chat message the participant reads it, and through the think-out-loud procedure states that they need to locate LZ Soontir. However, before shifting their attention to LZ Soontir, the participant checks back in with Gray Eagle's progress on the engagement of Wheeled Entity 152. The participant monitors and controls Gray Eagle's camera sensor until 44:30, at which time they then decide to pivot the Shadow's efforts from NAI 11 to instead begin addressing the need to recon LZ Soontir.

At 44:35 the user interface begins to lag in response to the participant's inputs, due to unforeseen artifacts of the synthetic world environment. Although verbally stating their task priorities, the participant struggles to work around the delayed system feedback. The system lag continues to be a challenge for the participant, and at 48:27 the experimenters choose to end the vignette. Thankfully this run through Vignette 4 was sufficiently long enough to capture the participant's chosen course of action and adaptation strategies to all of the scripted pop-up events.

At 48:27, the end of the vignette, the TOC asks the participant whether or not they advise the MEDEVAC operation to continue. Since the participant had neutralized the known threats in the battlespace, they confirmed with the TOC that the incoming MEDEVAC team could continue their mission, but advised of a known radar detection system to the North of the Ground Troops. After consulting with SMEs after the completion of this run, both SMEs observing the vignette agreed that this was a good call made by the participant.

Participant 2 Directability Summary. As discussed in Participant 7's summary on directability, two of the quadrants in Table 6 are of particular interest to the topic of supporting directability in interface design. The first gulf in directability emerged when *the participant had the authority/ability to direct actions but did not use it* — unfortunately to the detriment of their performance. As denoted in Quadrant III in Table 6, this type of directability gulf is considered a *missed opportunity* on behalf of the participant. There are two examples of this occurring in Participant 2's vignette. The first was the fact that Participant 2 did not launch two of their four available ALEs. This participant seemed to really maximize their use of assets *throughout* the battlespace, so it was possible the participant simply forgot about the two ALEs onboard the Grey Eagle UAV.

The second *missed opportunity* was Participant 2's subtle yet noteworthy difficulty interacting with the chat tile. More specifically, Participant 2 was unable to scroll to find past chat messages as they struggled with the design of the chat tile. Part of the struggle for this participant was that users first needed their mouse cursor *within* the tile before the mouse scroll wheel could be used to navigate through messages, and the scroll bar was hidden until hovered over with the cursor. The combination of these two hidden rules of usability (albeit common interaction styles seen in chat/messenger applications), paired with a high cognitive workload environment was enough to hinder Participant 2's ability to interact effectively with the chat tile. The larger consequence of this was a decreased situation awareness, as many chat messages influenced user activities and their prioritization of tasks.

The second noteworthy gulf in directability occurred when *the participant did not have adequate authority/ability to direct actions but needed the ability*. As denoted in Quadrant II in Table 6, this type of directability gulf resulted in an *intent violation or workaround behaviors* on behalf of the participant. One noteworthy example of this was observed when the participant relied on verbal commands to the Ownship pilot to direct the position of their aircraft for a threat engagement. The verbal commands were certainly a streamlined method of issuing directives to the Ownship, which could be a valuable ability to extend to the UAV control mechanisms in the future. This is an interesting area of potential future improvement in directability, since different tactics were able to be used by the more experienced participant aviators through verbal commands. As a note, Participant 2 was one of the more experienced aviators in the Round 2 Evaluation, with 3500 Total Flight hours (*M* = 2777, *SD* = 1966.19, Range: 295 – 5750), 3000 "Pilot in Command" hours (*M* = 1535, *SD* = 1322.78, Range: 0 – 3000), and 2000 Combat hours (*M* = 1435, *SD* = 1280.86, Range: 0 – 3400).

Discussion – Participant 2 Directability. It was interesting to observe multiple participants *verbally* commanding the Ownship aircraft (e.g., tasking, change flight maneuvers, change airspeed). This mode of communication may be simply quicker and/or less cognitively effortful. This could be due to the habitual use of radio communication in real-world operations, and/or the freed cognitive bandwidth of the underutilized audio – verbal – speech (i.e., perception – cognitive process – response) information processing resources. Most information in these evaluations was instead transmitted through the visual – spatial – manual (response) cognitive mechanisms (Wickens, 1980, 2002). Focusing future design efforts around enabling the user to use speech to direct the TECUMSA system could therefore help support what Rasmussen (1983) refers to as "skill-based behaviors" and increase overall system usability.

Directability Chapter Summary

Directability is the user's ability to take or direct actions to utilize system affordances. *Directability* in interface design is focused on the authority to specify an intention or command and to initiate action. This includes updating autonomy to the dynamic flow of resources, activities, and priorities, as well as making direct inputs (e.g., launching an ALE, confirming weapons launch).

In terms of sufficient directability, it is clear that the use of play calling was an efficient and effective way for a human operator to communicate high-level objectives to guide their autonomous teammate. This play calling approach also enabled the operator the ability to quickly transition between *monitoring* the flow of resources and *directing* the flow of resources. This ability to transition between passive and active resource management is valuable, as it allows the human to focus on other aspects of their job, but also step in if the automation reaches its competency boundaries (e.g., rare or unanticipated circumstances that the autonomy was not designed for).

In reviewing directability in Participant 7 and Participant 2's vignette performance, it was most helpful to consider two different directability gulfs: *Missed Opportunities,* or *Workaround / Intent Violation Behaviors* (as seen in Table 6). Missed opportunities were times when *the participant had the authority/ability to direct actions but did not use it* — unfortunately to the detriment of their performance. Examples included both participants failing to launch *all* of their available ALEs and thus having fewer assets out searching the battlespace for threats, as well as Participant 7's suboptimal tasking of the Fire-X to support the Ground Troops despite the aircraft's slow arrival time.

Workaround / intent violation behaviors on the other hand were times when *the participant did not have adequate authority/ability to direct actions but needed the ability*. Examples included the gap in coordination/directability between the participant and the autonomy in their criteria for selecting the "best" aircraft for a given play (e.g., selecting an available/untasked aircraft as opposed to the quickest aircraft to arrive), as well as when participants used verbal commands to direct the Ownship pilot as a substitute for the TECUMSA interface.

One of the more impactful gulfs in directability in the Round 2 TECUMSA evaluations was revealed by the differing weighting schemas used between the participants and the autonomous allocator agent when determining an aircraft's suitability for a play. Depending on the context of the play task, different criteria may be used to determine the "best" aircraft. For example, at what point is it worth pausing an ongoing play to free up a more desirable aircraft (e.g., able to arrive at the destination sooner)? To be able to more informatively direct acceptable aircraft selection operators needed the ability to coordinate the *priority* of a given play with their autonomous teammate. This is a parameter that in the future could be added to the Play Workbook tile as an optimizing criterion.

The reason for incorporating the *priority of play* as opposed to an *aircraft's availability* is that it has longer-lasting implications. *Aircraft availability* only matters at the time of play calling, while the *priority of the play* matters not only at the time of play calling, but also later in the mission should the system need to decide what play to temporarily suspend or cancel in lieu of a higher priority play. Play Workbook settings should therefore strive to not only communicate the user's intent at the moment but also carry meaningful information further into the mission (e.g., invest one button click now to avoid three button clicks later).

While improving directability in interface design is an admirable goal, the solutions for directability cannot be decoupled from observability. *This coupling between* *observability and directability, or perception-and-action, is what ultimately provides the user control over the system*. For example, designers can put the fanciest brake pedals in your vehicle, but it will not matter if you cannot see out the front window. In the next chapter on *controllability*, this interdependency between observability and directability will be discussed.

CHAPTER 7 - Control

Introduction

In the scope of this dissertation project, the human operator (AMC) was put in charge of a team of up to 7 UAVs as well as a manned aircraft. In this reorganized air assault operation, with this team composition, the *first major challenge* this AMC faces is *bandwidth.* A single operator − the AMC − cannot manually control the flight of 7 UAVs simultaneously. The operator, therefore, has to offload some portion of their responsibilities to autonomous systems. The *second major challenge* the AMC then faces is coordination with these subordinate autonomous systems. In both of these challenges, freeing bandwidth and coordinating goals, *the play calling approach provides a viable solution*.

Of significant importance is the *allocation of roles and activities* between the human and the autonomous systems. A variety of control systems and their suitability for this control problem will therefore be discussed, including manual control, shared control, and distributed supervisory control systems. Also discussed will be the aspects of play calling that provided *stable control* of the team of aircraft in this work domain.

Closed-Loop Systems

Mole et al. (2019) present a schematic diagram of the perceptual-motor control loop involved in safe driving, which has been adapted in Figure 35. More specifically, driving is a manual control task involving an active feedback loop, where perception continuously guides action, which results in output that interacts with the real-world and loops back to continue informing perception. For example, the driver is constantly observing their vehicle's distance from the vehicle driving in front of them. At a certain point, this distance will pass a personal comfortability threshold, which will guide their action of pressing the brake pedal. As the driver brakes, they are continuing to gather information to determine if the braking pressure is sufficient, which will have to be adjusted depending on the conditions of the roads, the tread on the tires, and other environmental factors. This perception-action cycle is thus referred to as a *feedback loop*.

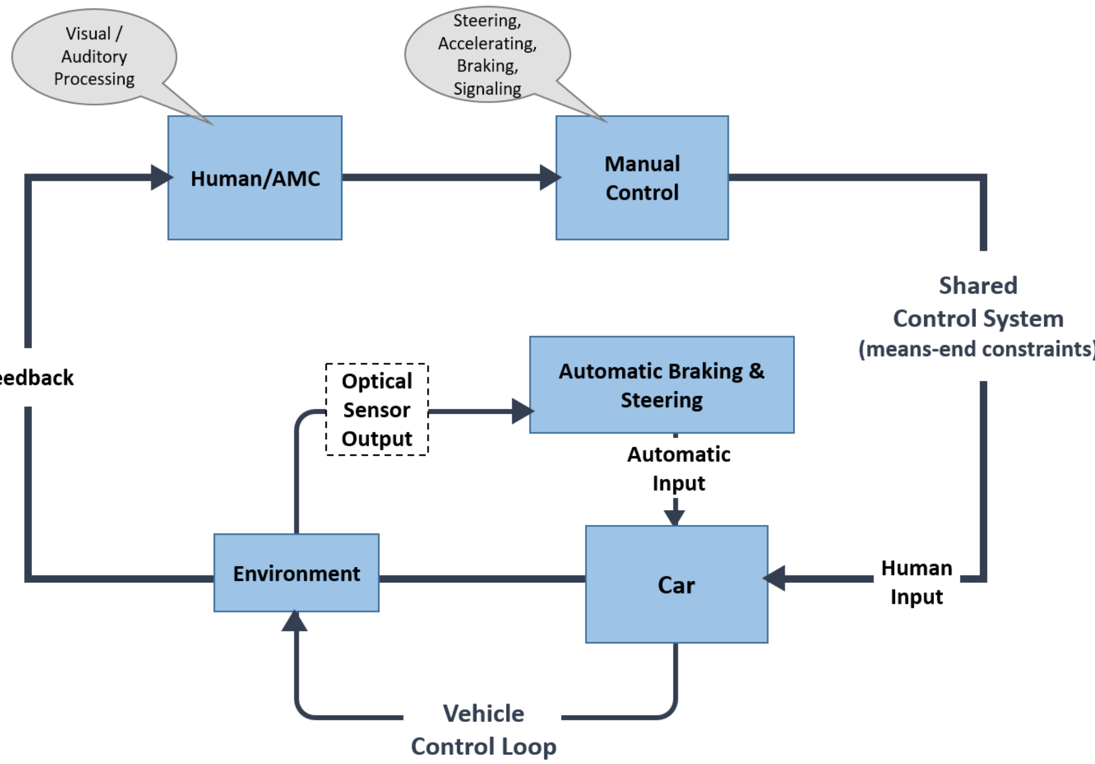


Figure 35. Shared Control System between Human and Smart Car Technologies such as collision-avoidance braking and lane-assist steering.

"Manual control of vehicles is made possible by the coordination of perceptual-motor behaviors (gaze and steering actions), where active feedback loops enable drivers to respond rapidly to ever-changing environments" (p.3, Mole et al., 2019). In that quote the authors are actually highlighting the original point of James Gibson (1979), that action-relevant information is both generated by and reciprocally used to regulate behavior. In Gibson's words "We must perceive in order to move, but we must also move in order to perceive" (p. 223, Gibson, 1979).

Control requires *both* sufficient observability and directability. A gulf in either ability will at best require compensation, and at worst will result in an unstable system. For example, if you close your eyes while driving, the car is no longer under control (even though you can actively change the trajectory and speed using the steering wheel and pedals) as you cannot observe your position on the road, the cars in front of you, and so forth. Without 'feedback' it is not possible to correct deviations from intentions (i.e., to control the vehicle). Similarly, controllability is also lost if you are unable to stop or slow down the vehicle due to a failed brake line (even though you can actively see the available distance to safely break shrinking). In this case, the feedback about a deviation from intention is observable, but there is no means to correct it (i.e., insufficient directability).

Controllability and observability do not have to be continuous, but they do require a sampling of information that is commensurate with the rate of potential changes (e.g., the bandwidth of the signal to be tracked). For example, drivers can safely look away from the road for only one to two seconds before performance is negatively impacted (Horrey & Wickens, 2007). In contrast, the need to check the fuel gauge in an aircraft may only be required a handful of times per hour during cruising flight.

A key factor in the stability of closed-loop systems is the amount of delay or lag (i.e., effective time delay) between the input and the resulting output of that system. Because information must be sampled at a rate adequate to respond to potential changes, delays or lags in feedback can easily lead to instabilities. Adjusting for delays in feedback therefore often requires operators to *predict* future states based on past inputs, thus reducing the relevancy of information that can be used to respond to environmental factors (Frank, 2018). For example, time delays can be a critical factor associated with pilot-induced oscillations in manual control tasks (Jagacinski, 1977).

Shared Control Systems

As Mole et al. (2019) discuss, another key factor in the stability of a closed-loop system is the dynamics introduced by multiple interdependent control loops, also referred to as a *distributed control system* (McCarthy, 2014). In more advanced automobiles, inputs such as the steering commands given to a car, are now the result of *multiple control loops simultaneously controlling the same system*. For example, when both human and automatic lane-assist technology give inputs to control the steering wheel (illustrated in Figure 35). This is known as *shared control* because control not only involves the human manually controlling their vehicle but also periodically receiving inputs from Smart Car technologies such as automatic braking and steering for lane correction.

The issue of "shared control" systems is that it introduces the need for precise coordination, collaboration, and trust. A collaboration/coordination problem can arise when the human and automation fight each other for control, and they have insufficient transparency into the goals and activities of one another (Miller, 2014). That is, a problem exists if the human perceives actions of the automation as *disturbances* that need to be corrected (e.g., lack in capabilities, lack of understanding). Additionally, trust issues are associated with the inappropriate calibration of trust (i.e., human's trust does not match the capabilities of the system) (Lee & See, 2004).

Of particular note to trust calibration is *overtrust*, where the human thinks that the automation is more capable than it really is (which may or may not be context-sensitive). Overtrust is particularly problematic when work activities have a high cost for failures (e.g., potentially fatal car crash). Therefore, the concern with this type of automation misuse (i.e. overtrust as opposed to distrust) is that the human does not attend to the controlling function because they assume the automation will be sufficient – also referred to as complacency (Parasuraman & Riley, 1997; Parasuraman, Molloy, & Singh, 1993). Complacency becomes a significant problem when the automation system fails, as the human does not have sufficient information (due to inattention) to recover control (Banks, Eriksson, O'Donoghue, & Stanton, 2018).

One of the main concerns that Mole et al. (2019) discuss is that the inclusion of automated vehicle technologies (e.g., lane-assist, adaptive cruise control) are actually changing the nature of the driving activity, due to their disruption to the typical manual control process that humans are accustomed to in vehicles. This includes difficult handoffs of control back to the human (e.g., poor operator situation awareness, insufficient response time), and the added workload that often accompanies the need to override autonomous systems when they reach their competency boundary of performance (e.g., lane-assist sensor loses track of the painted lines on the road while midway through a curve on the highway).

In shared control systems, although some portion of the human's manual control task is offloaded onto autonomous systems, it is only a portion of the total effort required for the human to maintain stable system control. It is thus important to note that in the context of air assault operations, where an AMC is in charge of a team of manned and unmanned aircraft, *shared control does not offload sufficient human workload to free the bandwidth required for stable manual control of multiple aircraft simultaneously.* Consider for example, that the automated technologies only take over control of the 7 UAVs sporadically, temporarily, and only to prevent serious route deviations or impending mid-air collisions. This control schema still leaves the human in primary control of 7 UAVs – an unstable solution at best, unfeasible at worst (Roth, Sushereba, Militello, Diiulio, & Ernst 2019). Hence the need to discuss *supervisory control* for this work domain.

Supervisory Control Systems

In more complex systems, there are often multiple layers of control loops needing to coordinate goals and activities (i.e., distributed control systems). For example, in military operations a hierarchy of command and control is used, where high-level objectives are issued by the commander, and her/his junior officers are then in charge of conducting the activities to actually achieve their commander's intent. Shattuck and Woods (2000) therefore describe this as a *distributed supervisory control system.* The authors write a useful excerpt explaining the dynamics of coordinating a commander's intent in a military operation:

> *"These [distributed supervisory control] systems are characterized by remote supervisors who work through multiple local actors to control a dynamic process. The agents can be separated by both space and time but still must coordinate their activities to achieve the goals of the system. Coordination normally occurs through the use of predetermined plans and procedures. However, these plans and procedures can be underspecified and brittle when a local actor is confronted with an unanticipated situation. In these instances, the local actor must adapt the plan in a manner*

> *consistent with the intent of the remote supervisor. Remote supervisors guide the adaptation by imparting their presence to local actors prior to controlling the process."* (p.1)

The future of warfighting is likely to continue requiring this command and control hierarchy that affords military commanders to set meaningful goals to guide subordinates' decentralized decision-making. In fact, the primary attribute of a distributed control system is its scalability due to the distribution of the control processing around nodes in the system. However, multiple layers of control loops mean that the right amount of coordination is required. Not only is coordination required between the commander and the subordinates to ensure goals and activities are aligned, but there also has to be coordination across subordinate groups to synchronize activities and ensure situation awareness is maintained (e.g., prevent friendly fire).

In real-world air assault missions, a top-down authority schema is commonly used. In other words, key decisions that determine the overall goal of an operation or mission are typically defined by a central agent. This control scheme largely hinges on a central figure, the AMC in this case, to constantly be informed of the changing dynamics of the battlespace (e.g., timing, location of forces both enemy and friendly, distribution of combat power). However, as Wickens, Hollands, Banbury, and Parasuraman (2016) point out, humans have limitations on the amount of information they can process since information sources must be sampled periodically, such as a pilot allocating attention through visual fixations to various cockpit instruments. With an informationally rich stimulus environment, critical events may be missed if operators do not select relevant stimuli to attend to at appropriate times.

In other words, there is only so much observability information that a single human can attend to, and thus there is only a finite amount of subordinate systems that they can reliably micromanage before succumbing to substantial lags in information processing and action-taking. That is, at a certain point the upper limits on human information processing will be reached, such as the bounds on visual attention and working memory capacity (Miller, 1956; Ma, Husain, & Bays, 2014; Franconeri, Alvarez, & Enns, 2007; Vidulich, Wickens, Tsang, & Flach, 2010). As the number of nested control loops increases (e.g., additional UAV's under the AMC's control) there is only one of two options: (1) A lag in processing or (2) Offloading of tasks/distributing control authority. *The AMC thus has to distribute authority to automation for continuous manual control tasks, and play calling is the interface for distributing this authority (i.e., directability).*

Using a *subsidiarity distribution of authority,* Sage & Cuppon (2001) advocate that power should be given to the lowest level decision-maker who has the necessary information to make the decision. The idea is to utilize a shift in authority from a top-down process that can suffer from lags in information transfer as well as insufficient information transfer, to a bottom-up process that will only suffer if subordinates are not properly trained or there is a misalignment in goals/priorities. In this organizational structure the main challenge is ensuring information is synchronized across decision-makers, so that deciders can coordinate a change in intent or resources, such as a significant loss in combat power.

It may help to clarify that the TECUMSA system used *both* distributed supervisory control, as well as subsidiarity distribution of authority. The AMC is the supervisor because they are still the primary control loop setting mission goals and parameters for what needs to be accomplished (i.e., calling plays such as “Recon Ingress Route Charlie”). However, part of advancing human-autonomy systems is to distribute decision-making authority across multiple agents. Therefore, the autonomous route planner was given the authority to re-route an aircraft without needing the AMC’s approval, to avoid certain high-risk circumstances including avoiding NFAs, known counter-air threats, and terrain collisions.

The use of distributed supervisory control (especially with the use of subsidiarity distribution of authority) does present a notable challenge. Specifically, when manual control tasks and certain decision-making authorities are offloaded to autonomous systems, *how does the supervisor (the AMC) coordinate goals and priorities with these subordinate autonomous systems?* This will be discussed next, along with an explanation for why the Play Calling and Play Workbook displays became such a critical design feature of the TECUMSA system.

Play Calling

As previously discussed, a single operator (the AMC) would not be able to dedicate the necessary bandwidth to pilot multiple aircraft simultaneously (e.g., monitoring all cockpit instruments and gauges, following a flight path, monitoring out the window for aerial obstacles). Instead, the human is better suited in a supervisory role, monitoring the battlespace from an “eagle-eye view” so they can make informed tactical decisions about the mission’s overall objectives. Subordinate autonomous systems can then be left in charge of deciding how to best achieve the supervisor’s goals.

The Play Calling and Play Workbook concepts are part of a necessary evolution in observability–directability coupling at the supervisory control level for this system. As previously discussed, the *play calling approach provides a predefined template for tasking that guides the autonomous agent's behaviors sufficiently to achieve the goal, but with enough degrees of freedom to optimize and adapt as needed* (Miller et al., 2004). For example, the AMC can call a Route Inspect play, but if somewhere along the route an NFA pops up, the autonomous route planner can plan around the obstacle.

Using TECUMSA, the supervisor (AMC) was able to coordinate goals and priorities with their subordinate autonomous systems because play calling allowed the AMC the flexibility to specify different details of intent as the battlespace changed and evolved. This effectively left automatons to handle the details at the manual control level (e.g., adjusting heading, airspeed) required to actually carry out the play. For example, when calling a play to task an aircraft to search an area of interest, the operator had decreased observability into the exact scan pattern that would be used by the camera sensor onboard the aircraft. However, it is important to note that because the operator was able to *directly specify the intent* of the area search (e.g., indicating the priority location of interest), it was no longer necessary for the operator to continually monitor or adjust the camera's scan pattern since the *autonomy's goals were aligned with their own*. Play calling, therefore, enhances the observability-directability coupling at the supervisory level, so that the AMC can precisely specify intent to the automation and provide immediate feedback to any deviations from that intent.

Consider the *tight perception-action coupling* observed in the Round 2 Evaluation, when the AMC was sitting inside the same cockpit as the human Ownship pilot. The AMC could clearly observe the aircraft maneuvers – and thus quickly communicate needed adjustments if the pilot's maneuvers were not consistent with the AMC's intentions. Although being co-located in the Ownship cockpit provided the AMC with thorough observability and directability into the behaviors of that aircraft (e.g., small airspeed adjustments, altitude adjustments), it comes at the expense of attention for other assets. More specifically, it would be difficult or impossible to allocate this same level of attention simultaneously across several aircraft. In contrast, commanding remote aircraft *through play calling*, despite it being a looser coupling between observability and directability, offers a solution to the bandwidth limitations of a single operator.

So the question then becomes, what afforded a stable system in these Round 2 vignettes? More specifically, *how did 'stability or instability' emerge as a result of this looser coupling of observability and directability*? The next two sections will explore observability − directability coupling *under the unique dynamics of a distributed supervisory control system*. This includes illustrating the control mechanisms in the TECUMSA system. Following that, will be a discussion of how play calling helped the AMC achieve stable control, as well as where it fell short of achieving stable control/performance.

Distributed Supervisory Control System: Diagram

Before answering the question "*What enables stability in this distributed supervisory control system,"* it will be most helpful to first illustrate the control mechanisms that existed in the TECUMSA system. Figure 36 illustrates TECUMSA's multiple control loops that are behind the shorthand phrase "human-autonomy teaming." The *outer loop* is the AMC's *supervisory control*. This is where the AMC issues tasks to aircraft through the *Play Calling tile*, and coordinates goals/intentions with the autonomous systems (e.g., allocator agent, autonomous route planner) through the *Play Workbook tile* (described in detail in the *TECUMSA: Play Calling Process* section). The *inner loop* is the *vehicle control*, requiring direct manual control of the aircraft and their payloads (e.g., adjusting heading, maintaining altitude, steering camera sensor).

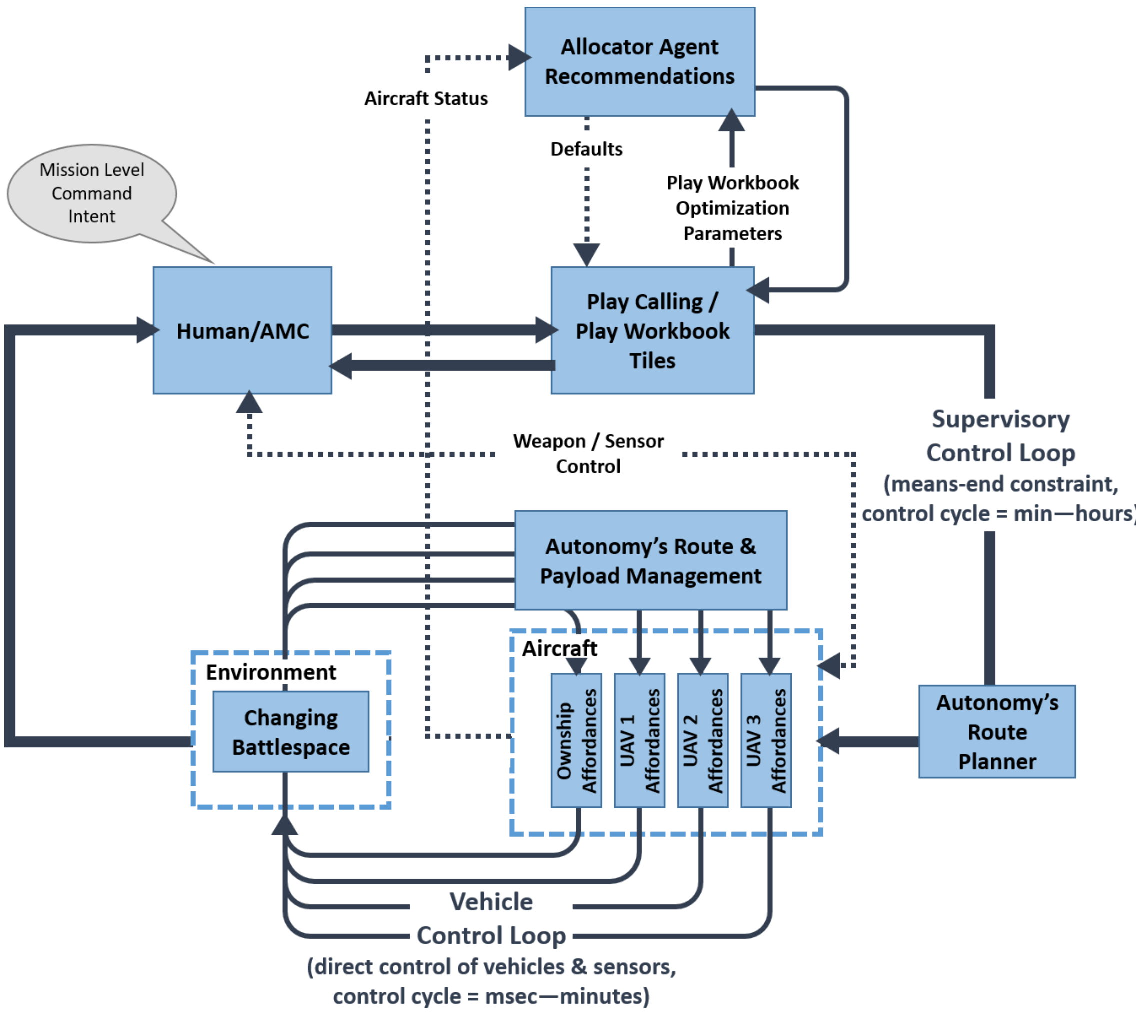


Figure 36. Multi-Control Loops in the TECUMSA System, Including a Supervisory Control Loop and Vehicle Control Loop.

Arrows indicate the general direction of information flow. The thick black lines indicate the (outer) supervisory control loop, and the thinner black lines indicate the (inner) vehicle control loop. This control scheme is not a simple nesting of control loops (as seen in Figure 35), rather *there is coordination/distribution of authority between the inner and outer loops*. Intermediate couplings between the inner and outer loops are represented by the black dashed lines.

It is important to note that *system stability emerges from the coupling of these two loops.* As you can see in Figure 36, this coupling is achieved by mapping the AMC's intentions onto the Play Calling / Play Workbook tiles. This mapping is first transferred to the allocator agent through the Play Workbook's settings (e.g., AMC clicks a button to optimize for *stealth* when choosing an aircraft for the play), so that the aircraft selection criteria can be aligned between the human and allocator agent. The AMC's inputs, mediated by Play Calling / Play Workbook tiles, provide enough details so that the Autonomy's Route Planner can then generate a route plan that the autonomous vehicle control systems will follow (e.g., considering max bank angles of each aircraft, aircraft's current location). This mapping is then intermittently updated – during play calling or play editing – so that the AMC and autonomous systems can remain well-coordinated.

The inner vehicle control loop then uses the output from the Autonomy's Route Planner as guidance for waypoint following and aircraft control. Feedback from the changing environment informs not only this continuous vehicle control loop (e.g., rerouting an aircraft mid-flight to avoid an obstacle) but also the goals and objectives of the AMC (e.g., calling a new play after a threat is detected to the north).

Intermediate couplings (black dashed lines in Figure 36) include the AMC's ability to override autonomous control, such as the camera sensors should the operator need to take over manual control (i.e., pan, tilt, zoom sensor). Additionally, an intermediate coupling exists between the vehicle control loop and the allocator agent, where up-to-date aircraft state information informs the allocator agent's aircraft recommendations. Aircraft state information includes the current play status of aircraft (e.g., available for tasking versus already busy on a play), as well as the aircraft's current location in the battlespace or distance from the objective.

Part of the elegance in the distributed supervisory control schema for this problem domain is highlighted by the different temporal requirements of the inner and outer control loops. One key characteristic of the inner (vehicle control) loop shown in Figure 36 is that for flight control actions to remain well-calibrated to the changing environment, control has to operate at a sufficiently high frequency – inputs issued on the order of milliseconds to seconds. On the other hand, the outer (supervisory control) loop is able to stay well-calibrated when control operates at a lower frequency – inputs issued on the order of minutes to even hours. The circumstances here dictate that the human can only stably direct multiple aircraft at the supervisory level, which frees valuable cognitive bandwidth that can then be used to make important commander decisions.

It is worth noting that the figure is simplified to only have three UAVs shown in the inner control loop, but in the Round 2 Evaluation this number was actually up to seven UAVs. It is also important to clarify that for the Ownship specifically, a member of the evaluation team served as the pilot, and only that aircraft's payload (e.g., camera sensor) was autonomously controlled.

In essence, directability is about how much authority you are willing to allocate to subordinates/automation, versus retain (i.e., AMC is in sole control). The ability to allocate authority to teammates is the key in supervisory control since it is no longer a single control loop with one actor perceiving and acting. In a complex socio-technical system, inputs such as the commands given to a team of aircraft, are now the result of multiple control loops coordinating together.

Ultimately, there are design solutions that will lead to stability and solutions that will lead to instability. The degree of directability given should be appropriate to the AMC's level of observability. For example, directability that requires higher rates of observability, such as manually steering an aircraft, is better left to automation since it requires such a high rate of information sampling and input correction to maintain stable control of the aircraft. Plays are therefore a valuable interface for the AMC to distribute authority by leaving degrees of freedom for the automation to resolve the inner vehicle control loop (e.g., re-route around NFA, adjust heading).

Distributed Supervisory Control System: Finding Stability

If the question is "*What enables stability in this distributed supervisory control system,"* the answer is *"Stability emerges from the coupling of the inner and outer control loops"* (as seen in Figure 36). More specifically, the supervisory loop must be well-tuned to the vehicle control loop and vice versa. Predominantly, the active coordination of intent occurs during the AMC's interactions with the Play Calling / Play Workbook interface. This is where, as supervisory controller, the AMC's goals and priorities can be aligned with the supporting autonomous systems that are in charge of the inner (vehicle control) loop. Because of this relationship, the Play Calling / Play Workbook designs are essential to enabling the AMC the ability to specify their intent to the autonomous systems.

The Round 2 Evaluation revealed cases where the coupling between the AMC's input and the resulting output of the autonomy were not as aligned as they needed to be, thus causing an “instability” in system control (as discussed in Chapter 5 and Chapter 6). Now that Figure 36 has been discussed, it will be helpful to review these *instabilities* at their respective locations within this control schema. Additionally, there are notable points of *stability* within this distributed supervisory control system that will also be discussed.

Control Instability #1: Missing Feedback from Autonomy Route Planner to Play Calling / Play Workbook Tiles ─ Estimated Time to Objective Not Provided

Note the bidirectional information flow between the AMC and the Play Calling / Play Workbook tile as depicted in Figure 36. Here a notable HMI design improvement opportunity was revealed in the Round 2 Evaluation. Missing information, specifically the estimated time it would take an aircraft to arrive at its destination (Estimated Time Enroute − ETE), would have significantly affected the AMC’s decision-making. In one case, Participant 2 realized a few minutes after tasking one of their slowest flying UAVs that it was not going to arrive at its destination in a practical amount of time. Although this participant caught the problem relatively quickly, the consequences of not having this ETE information can be far more significant (as Participant 7 found out).

So although the human/AMC could see the planned route in the map before accepting the play, there was still not sufficient transparency into the route planner’s estimated time for the aircraft to travel to its destination. Adequate feedback from the Autonomous Route Planner to the human/AMC would have thus tightened the coupling between the supervisory control loop and the vehicle control loop.

Control Instability #2: Additional Coordination of Constraints needed from Allocator Agent and Play Calling / Play Workbook Tile

The most important aspect of the Play Workbook tile is its ability to provide the AMC with flexibility in their specificity of intent. What is meant by *specificity of intent* is that the required user inputs provide enough specificity to appropriately guide the autonomy's aircraft allocation and routing choices, but preserves enough degrees of freedom for the automation to resolve the manual control tasks in the inner loop and respond to pop-up events (e.g., routing around an obstacle). Consider a real-world example when you enter a destination into your car's navigation system. You have the ability to avoid ferries, avoid highways, avoid tolls, avoid unpaved roads, and prefer fuel-efficient routes. However, there are still parameters that might be missing in certain circumstances, such as the need to avoid low bridges when driving your belongings cross country in a moving truck.

Similarly, the Play Workbook lacked sufficient ability for the AMC to communicate their intent across *all* circumstances. One such example of this insufficient directability occurred when the participants needed to support the Ground Troops under ambush. Many of the operators needed the ability to mark a play as the *most* time-sensitive and high-priority play up to that point in the mission, their goal was to get *any* armed aircraft over to the troops in the least amount of time.

This particular event was extremely valuable, as it revealed that there was a misalignment between the allocator agent and AMC regarding their criteria for what makes an aircraft "best" for a play. This misalignment was magnified by the fact that certain constraints of the allocator agent's reasoning were hidden from the AMC (i.e., preference for available/untasked aircraft). This made the coordination of goals between the AMC and allocator agent much more difficult, since even after a period of interacting with the system the AMCs were still not always able to pick up on this hidden preference of the allocator agent.

As you can see in Figure 36, inputs from the AMC are mediated by the optimization parameter options available in the Play Workbook. Notably, there actually was a location in the interface where the participant could see plans, independent of an aircraft's availability (see Quick Swap tile discussed in the *Play Workbook: Quick Swap Tile* section). The problem was that the Quick Swap tile was not *immediately* accessible (i.e., it required opening a separate tile from within the Play Workbook). This highlights the importance of *coupling* both observability with directability. In this case, the operator needed both observability/transparency into the hidden preferences of the allocator agent, as well as the ability to direct the allocator agent so it could align its priorities to that of the AMC.

Control Stability #1*: *Coordination of Control between AMC and Autonomy's Route & Payload Management

In Figure 36 intermediate couplings between the inner and outer loops are represented by the black dashed lines. These are connections where the human always had ultimate authority to take over control (i.e., weapons, camera sensors, launching ALEs). For example, the AMC was able to take over manual control of the camera sensor on any aircraft (i.e., pan, tilt, zoom) by simply making control inputs in the Vehicle Dashboard. Sensor control would only be taken back by the autonomy if it was required to complete a play and the human was not actively steering it. Observations from the evaluations support that this enabled a graceful swapping of sensor control between the human and the autonomy, thus enabling a tighter coupling between the inner and outer control loops (seen in Figure 36).

Another intermediate coupling between the inner and outer control loops was the human's authority over weapon control. The distribution of authority during a kinetic threat engagement can be seen in steps 10 through 15 below, pulled from Chapter 6 (*Participant 7's Detailed Narration - Directability*). Due to the high consequences of engaging threats, there was a deliberate splitting of tasks between the human and the autonomy during an engage play. Of note, the ultimate authority to end/delay the engagement or fire the weapon was allocated to the human (as represented by the black dashed line in Figure 36).

Handovers and *takeovers* of control between humans and automation are challenging. Drexler, Takacs, Nagy, and Haidegger (2019) review the handover process of autonomous vehicles, where they highlight the risk of incomplete information during the handoff, as well as inadequate response time. However, the task allocation or division of responsibilities between the AMC and autonomy in TECUMSA during (kinetic) engagement plays appeared to be a stable distribution of control based on experimenter observations and feedback from the Round 2 Evaluation. The successful handoffs are thought to also be in part due to aviator training. Aviators verified that engagements are of very high importance and the acceptable margin for error is small, which likely influenced their tendency to focus their attention and maintain situational awareness on the engagement task. Refer back to Table 5 to see the typical workflow from start-to-end of an engagement play in TECUMSA, as well as the task allocation between the human (AMC) and the autonomy.

Control Stability #2: Autonomy's Route & Payload Management

Overall, the autonomous system's control of the aircraft and payloads during the simulated flight was stable (e.g., waypoint following, airspeed corrections). Routes were successfully generated, all aircraft followed these routes, and no aircraft were flown into the side of a mountain! Additionally, the route planner successfully avoided flying aircraft too close to certain entities once the AMC marked them as a high-risk threat. This avoidance behavior was even aircraft specific, as some aircraft in the simulation were more at risk of being shot down depending on how close they flew to a threat.

However, recall that in the supervisory control loop high-level guidance is issued through the Play Calling / Play Workbook tile and then carried out by the subordinate autonomous systems. This means that so long as the AMC's intent was able to be meaningfully translated through play calling, manual control of the Autonomy's Route & Payload Management should be successful (e.g., a Point Inspect Play for Building 14 means a route needs to be generated that is the shortest possible distance for an aircraft with a camera sensor to travel to Building 14). The requirements for stability are built into the autonomous control systems and reflect the appropriate distribution of authority built into the play calling logic.

Summary

The *first major challenge* the AMC faces is *bandwidth.* A single operator – the AMC − cannot manually control the flight of 7 UAVs simultaneously. The operator, therefore, has to offload some portion of their responsibilities. A significant amount of time can be saved if computer-generated routes can be used, rather than hand-plotted routes. Furthermore, the task of actually following the prescribed flight route would demand near-continuous operator inputs to ensure all 7 UAVs correctly followed their planned route (e.g., adjusting heading and airspeed, monitoring altitude and possible nearby obstacles). However, this too can be offloaded so that the AMC does not have to be consumed by the task of manually controlling each UAV.

The *second major challenge* an AMC faces is *coordination* with subordinate autonomous systems. Although notable bandwidth is freed by offloading route planning and manual control of aircraft to automated computer systems, a new challenge emerges for the AMC as they must now maintain coordination between their goals and the goals of the automated systems. An HMI that enables adequate specificity of the AMC's intent is therefore required.

In both of these challenges, freeing bandwidth and coordinating goals, *the play calling approach provides a viable solution*. First, play calling offers the *benefits of subsidiarity distribution of authority*, where intelligent autonomous agents can adaptively make route planning and flight control adjustments without the need to task the AMC. Additionally, play calling offers a method of *supervisory control* where the AMC is able to disseminate operational guidelines / goals with enough detail to guide the intelligent agent's decisions, but also enough degrees of freedom to permit adaptations during execution – a concept referred to here as the *necessary specificity of intent*.

With the Play Calling / Play Workbook tiles being the central hub for mediating the AMC's intentions into actionable plans by the autonomy, it was incredibly important to note instances where the design did not provide sufficient observability or directability.

Certain circumstances in the Round 2 Evaluation revealed opportunities to improve the AMC's supervisory control. One such case was the misalignment in priorities/values between the human and allocator agent when determining the desirability of an aircraft for a play (e.g., availability status versus time to reach the destination). Another notable instability was revealed when the AMC did not have sufficient observability into the autonomy's estimated time for how long it would take for an aircraft to reach its destination.

Most importantly, however, the results of the Round 2 Evaluation revealed that stability emerges from the coupling of the outer (supervisory control) loop and the inner (vehicle control) loop. That is, observability-directability coupling must be considered not only within each control loop but *across* control loops as well. Although these control loops function separately, *it is the coordination of intent, priorities, and activities between the supervisory control loop and the vehicle control loop that ultimately produces a stable system*

CHAPTER 8 - Discussion

Problem Space

In the scope of this research, two rounds of evaluations took place where participants (all Army aviators) acted as the AMC of an air assault mission. During real-world air assault missions, the AMC is the overall air mission leader and designated representative of the aviation unit for the mission (Army Training Circular No. 3-04.11, 2018). Just as in real-world operations, participants were delegated with decision-making authority (e.g., tactics, priorities, limitations) over their team of manned and unmanned aviation assets to successfully complete the objectives of an air assault mission within a synthetic task environment.

Participants used the TECUMSA system to effectively monitor and manage their available team of aircraft and resources (e.g., missiles, rockets, camera sensors) to accomplish tasks that are common in Army air assault operations. More specifically, in these evaluations, the AMC was responsible for accomplishing a series of reconnaissance, surveillance, and threat neutralization tasks in a hostile, dynamic, and unpredictable battlespace. One major challenge investigated in this research is the changed work requirements of a single AMC being in command and control of up to 7 UAVs, as well as their manned aircraft (Ownship).

As the number of nested control loops increases (e.g., additional camera sensors under the AMC's control, additional UAVs under the AMC's control), the only stable option is offloading of tasks or distributing control authority. Even two camera sensors would be difficult for a single AMC to concurrently manually control (e.g., simultaneous visual searching, panning, and zooming to track two different moving objects). In the scope of this research distributed control authority was explored, where *the AMC distributed authority to automation for continuous manual control tasks, using Play Calling as the interface for distributing this authority (i.e., directability)*. The stability of this solution will be discussed in the following sections.

Coordination in Distributed Control Systems

Sage & Cuppon (2001) advocate that power should be given to the lowest level decision-maker who has the necessary information to make the decision. The idea is to utilize a shift in authority from a top-down process that can suffer from lags in information transfer as well as insufficient information transfer, to a bottom-up process that will only suffer if subordinates are not properly trained or there is a misalignment in goals/priorities. Aptly named, the authors refer to this as a *subsidiarity distribution of authority*. In this organizational structure, particularly one with human-autonomy teaming, the main challenge is ensuring information and priorities are synchronized across decision-makers so that deciders can coordinate a meaningful change in intent or resources (e.g., a significant loss in combat power).

Use of Play Calling

The play calling approach applies the subsidiarity distribution of authority to enable the AMC to specify command intent, but leave the subordinate autonomous systems to handle many of the specific decisions about how to achieve that intent. More specifically, the TECUMSA user was presented with a bank of high-level tasking options, or “plays,” to quickly communicate to the autonomy a higher-order objective (e.g., "I want to look somewhere, "I want to follow something", “I want to kinetically engage something”). With that information, the autonomy was able to make a recommendation to the human for what aircraft it thought would be best to achieve that goal, and plan a flight route for the suggested aircraft. For example, by calling a Parallel Search play the operator directed the autonomy to conduct reconnaissance on a larger area, at which point the autonomy could then recommend an aircraft such as the Grey Eagle UAV, and then plan a suggested route to utilize the aircraft’s far-range camera sensor to quickly scan the large area of interest. This control organization in TECUMSA was designed so that the human was still the high-level decider of goals and objectives by tasking aircraft through “play calling” (supervisory control), while the autonomy’s role was deciding the most efficient way to accomplish the task at hand and adapt the route as necessary during play execution (subsidiarity distribution of authority).

Theoretical Implications of Play Calling

Coordination and collaboration become the main challenges for the stability of system control using the play calling approach, because play calling was the sole medium for the human to specify their goals, priorities, and desired activities to the subordinate autonomous systems. This research focused on designing for humans to effectively coordinate and collaborate with autonomous agents, where observability and directability were two core design principles. Simply put, *observability* is about the user having the ability to understand and explore the solutions generated by the autonomy. *Directability* is about the user’s ability to choose the right intent (goal) and to specify it clearly to the effectors (the aircraft), while utilizing autonomous agents to generate meaningful solutions based on this intent. In other words, observability was about *seeing* the solution space, whereas directability was about the *authority/ability to guide, adjust, and select* the appropriate solution.

Practical Implications of Play Calling

The following three sections will cover a few of the practical implications, including benefits and challenges, of the play calling approach for the TECUMSA system.

Benefit: Flexible Specificity of Intent

Play calling facilitated goal coordination between the AMC and the subordinate autonomous agents. The Play Calling / Play Workbook interface gave the AMC flexibility in their *specificity of intent*, where plays provided a predefined template for tasking that guided the autonomous agent's behaviors sufficiently enough to achieve the goal (e.g., recommend aircraft, plan route), but with enough degrees of freedom to optimize and adapt as needed (e.g., avoid flying through NFAs, avoid new counter-air threats) (Miller et al., 2004).

Furthermore, the user could adapt a play based on their higher-order objectives, by inputting optimizations (e.g., "speed of task completion," "combat power"), which would allow the autonomy to make a more informed recommendation based on the *situational values and priorities* of the user. The design of the Play Workbook was able to subtly shape the behavior of the user in a way that encouraged interactions at a more efficient and effective level of control (i.e., supervisory control as opposed to manual control). The design of this interface component allowed the user to both clearly *observe* the possibilities for tasking aircraft and quickly *direct* the desired solution without having to take an excessive number of actions.

Benefit: Freed Bandwidth

Play calling was an incredibly valuable control mechanism in this distributed supervisory control system. Plays enabled a single operator to stay in command of a team of up to seven unmanned aircraft simultaneously. Play calling also enabled the operator the ability to quickly transition between *monitoring* the flow of resources and *directing* the flow of resources. The ability to quickly direct or manage aircraft tasking was valuable as it allowed the user/AMC more cognitive bandwidth to focus on other aspects of their job (e.g., maintaining situation awareness). In times when the AMC was particularly busy, the ability to quickly get a task underway was critical so that they could return to monitoring the evolving battlespace. For example, enemy missile launches were an especially cognitively demanding and stressful time for the AMC. The AMC's ability to maintain situation awareness hinged on their ability to monitor when and where the missiles were being launched. Rather than having to spend excessive time to direct aircraft activities, play calling enabled the AMC to allocate attention flexibly across the seven aircraft as well as the dynamic battlespace.

Challenge: Precise Coordination of Goals

Coordination between humans and autonomous agents requires precise alignment of goals, priorities, and values. Unforeseen scenarios can reveal gaps or misalignments in the coordination abilities between the human (AMC) and the autonomy, which makes system evaluations in synthetic task environments so valuable. One of these unexpected scenarios included the pop-up event in Round 2 where Ground Troops suddenly were ambushed and required the AMC to provide immediate armed assistance. This event revealed that there was a misalignment between the autonomous allocator agent and the

AMC regarding their criteria for what makes an aircraft “best” for a play. Generally, the allocator agent’s selection criteria for “best,” which included prioritizing the use of aircraft that were un-tasked, worked well for achieving the AMC’s goals (i.e., avoiding having to pause or cancel ongoing plays if there are already suitable aircraft available).

However, in this particular case of “troops under fire,” participants generally felt that the most important factor in deciding the best aircraft was merely finding the *armed* aircraft with the *shortest estimated flight time*. Participants were typically willing to suspend any ongoing task to maximally reduce the time needed to arrive and support the Ground Troops. Participant 7’s late realization that Fire-X was not going to arrive in an acceptable amount of time to the Ground Troops indicates that they were unable to effectively coordinate their goals with the autonomous allocator agent to guide its recommended aircraft tasking solutions to assist the Ground Troops.

Interestingly, there actually was an option in the interface that enabled the user to include or “Show Busy Aircraft” as part of the allocator agent’s recommended solution (see Quick Swap tile discussed in the *Play Workbook: Quick Swap Tile* section). However, the insufficient human-autonomy coordination of priorities (AMC’s goal to get *any* armed aircraft to Ground Troops as soon as possible), paired with insufficient observability (lengthy flight time from the available Fire-X, versus immediate support from busy Grey Eagle), meant the user did not even realize they needed to use this feature of the interface.

This lesson highlights the importance of a tight *coupling* between observability and directability. In this case, the operator not only needed observability or transparency into the hidden preferences of the allocator agent (i.e., preference for available/untasked aircraft) but also the ability to use that information to then properly direct the allocator agent so it could align its priorities to that of the AMC (i.e., get *any* armed aircraft to the Ground Troops as soon as possible). In TECUMSA, the AMC's intentions were mediated by the options available to them in the Play Calling / Play Workbook displays. It was therefore invaluable for system robustness to evaluate the user's ability to communicate their intentions *across a variety of circumstances.*

Theoretical Implications: Observability & Directability in Design

Observability and directability were also core design principles outside of the need for human-autonomy teaming. *Interfaces must be designed to indicate the meaningful critical action boundaries relative to the operator's goal* (Hutchins, Hollan, and Norman, 1985; Norman, 1986). This includes designing observability into the possibilities of action, consequences of action, status feedback, as well as transparency into the behaviors and capabilities of the autonomous agents (Sarter et al., 1997; Miller, 2014).

The user interface design acts as the medium for users to interact with the real-world affordances. This means that while affordances are the possibilities, directability (through the interface) is how you choose to navigate or use the possibilities. So, although the AMC is not able to change an aircraft's max airspeed (an affordance of the aircraft), they are able to choose or direct whether or not to fly the aircraft at that max speed. Ideally, there is also adequate transparency, or observability into the consequences of making the choice to fly at max airspeed (e.g., quicker arrival to destination, but increased fuel consumption and thus decreased range of flight).

Retrospective

The following two sections will cover lessons learned from the Round 1 Evaluation as well as a brief discussion on the importance of having SME involvement in this complex interface design challenge.

Lessons from Round 1

Over the three years of this research, countless insights were gathered to improve the human-autonomy teaming and interface design. Although the focus of this dissertation centered on the Round 2 Evaluation, there were valuable lessons also learned from the Round 1 Evaluation that were implemented to improve the TECUMSA system. During the Round 1 Evaluation, it became clear that the interface design needed to streamline the efficiency of tasks involving finding, identifying, and engaging threats, which ultimately shaped an AMC's overall mission efficacy. This was particularly apparent when participants were observed making frequent switches back-and-forth between the two monitors; users commonly switched focus between the video feeds in the Vehicle Dashboards on the left monitor, and the Map and Play Calling displays on the right monitor.

This observation led to design changes that made information and inputs more accessible *at the point of need,* also known as increasing visual momentum (Woods, 1984). For example, rather than make participants solely use the map's radial menu to reclassify entities as hostile, in Round 2 the ability to reclassify an entity was added to the Vehicle Dashboard as well. This was because participants' attention would already be in the video feeds of the Vehicle Dashboards when they identify an entity as hostile. To streamline this reclassification process, a list of all entities in that sensor's current field of view was placed just to the left of the video feed with right-click interaction to quickly reclassify entities (thus minimizing the back-and-forth between monitors).

SME Involvement

I cannot fully capture the importance of continued inputs from SMEs throughout the design and evaluation process. Not only were SMEs active contributors during design ideation and generation, but they were also present during both rounds of evaluations. SME feedback and insights throughout the iterative design and evaluation process was integral to the quality and ecological robustness of the TECUMSA system. To offer just one of countless examples of the value they added, there was an early design for the "Fire Now" window (mentioned earlier in this paper) that had the user confirming weapons release through a pop-up window that would appear in the map. After showing the SMEs on the team the design, they pointed out that the AMC's attention is predominately going to be in the video feeds of the Vehicle Dashboards since that is how they maintain eyes-on the target (i.e., PID) before confirming weapons release. The current placement of the "Fire Now" window in the Vehicle Dashboards (discussed earlier) is the result of their simple yet critical insight regarding the typical threat engagement workflow.

Limitations

This research had multiple limitations. First, this study was limited to a small number of participants due to budgetary, time, and availability constraints. Having a relatively small number of participants in the evaluations means that conclusions and observations may not be generalizable.

When designing a system for a particular user, such as an Army helicopter pilot, it is incredibly valuable to test and evaluate the system using that specific user population.

For the two rounds of evaluations of the TECUMSA system, Army helicopter pilots were brought in as participants. While there are sizable advantages to using active duty pilots (e.g., recency of tactical knowledge), there are also negatives. One of the disadvantages is that their time is extremely valuable, which meant we were limited to only having a single day to train participants on TECUMSA. This time limitation meant participants were not able to truly master their use of the system. In fact, certain results, such as participants not using the Quick Swap tile to choose aircraft, are likely tied to a lack of training.

It is also worth mentioning there are notable differences between a simulated versus real-world air assault mission. There is far more information being transmitted in a real-world operation. For example, pilots frequently have to listen to several radio channels simultaneously, as well as monitor and reply to multiple text-based chat discussions. The results in this research are therefore limited by the difference in information flow. For example, the effectiveness of the audio alerts may be significantly reduced with additional audio distractions such as those experienced in the real-world. Additionally, the vibration in a helicopter poses a unique challenge for touchscreen interfaces and computer mouse control-ability, which could not be tested in this stationary simulation environment.

Future Research

As the number of aircraft under the AMC’s command increases, the number of alternative options to achieve the play’s goal grows exponentially. When users are faced with a set of tradeoffs, providing summary information of the largest gain and loss can simplify decision-making. In future research, it would be valuable to hone the decision-making support offered by the Quick Swap tile. More specifically, to supplement the parallel coordinates plot there is information that shows the user the “biggest advantage” and “most notable loss” if they choose an alternative plan to the one recommended by the autonomy. In the future, course of action comparisons could be refined to provide more concise and easily interpretable values to indicate not just the largest gain and loss of a given plan, but also the magnitude of each. Additionally, other meaningful details such as the consequences of choosing a particular course of action may also support AMC decision making and increase transparency into autonomous agent reasoning (e.g., choosing to optimize on speed will not allow for any additional reconnaissance during ingress due to the location of the flight path respective to nearby key terrain).

Finally, there is always room to improve the domain-specific intelligence of autonomous systems. As discussed previously, an ability for the AMC to specify the priority of a given play could be used to increase the scope of decision making on behalf of the autonomy, to include more dynamic aircraft re-tasking strategies mid-execution, as well as active recommendations for more optimal mission completion if resources/capabilities change in a meaningful way (e.g., autonomy suggests reconning a nearby point of interest during an aircraft’s ingress flight to another task).

REFERENCES

Apker, T. B., Johnson, B., & Humphrey, L. (2016). LTL templates for play-calling supervisory control. In AIAA Infotech@ Aerospace (p. 0917).

Army Field Manual. (2015, March). Airborne and Air Assault Operations (FM 3-99). https://armypubs.army.mil/ProductMaps/PubForm/Details.aspx?PUB_ID=105021

Army Field Manual. (2020, April). Army Aviation (FM 3-04). https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN21797_FM_3-04_FINAL_WEB_wfix.pdf

Army Training Circular. (2018, September). Commander's Aviation Training and Standardization Program (TC 3-04.11). https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN14459_TC%203-04x11%20C1%20INCL%20FINAL%20WEB.pdf

Banks, V. A., Eriksson, A., O'Donoghue, J., & Stanton, N. A. (2018). Is partially automated driving a bad idea? Observations from an on-road study. Applied ergonomics, 68, 138-145.

Bennett, K. B. & Flach, J. M. (1992). Graphical displays: implications for divided attention, focused attention, and problem solving. Human Factors, 34, 513 – 533.

Bennett, K.B., Posey, S.M., Shattuck, L.G. (2008). Ecological interface design for military command and control. Journal of Cognitive Engineering and Decision Making 2 (4), 349–385.

Bradley, J. (2019, May 23) [Picture of Soldiers using the Future Open Rotorcraft Cockpit Environment, or FORCE, Simulator]. *Aviation, Missile Center simulator supports Army readiness*. Retrieved from https://www.army.mil/article/222253/aviation_missile_center_simulator_supports_army_readiness

Calhoun, G., Draper, M., Miller, C., Ruff, H., Breeden, C., & Hamell, J. (2013, September). Adaptable automation interface for multi-unmanned aerial systems control: preliminary usability evaluation. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 57, No. 1, pp. 26-30). Sage CA: SAGE Publications.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. Cognitive science, 37(2), 255-285.

Cooke, N. J., Rivera, K., Shope, S. M., & Caukwell, S. (1999, September). A synthetic task environment for team cognition research. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 43, No. 3, pp. 303-308). Sage CA: Los Angeles, CA: SAGE Publications.

Cummings, M. M. (2014). Man versus machine or man+ machine?. IEEE Intelligent Systems, 29(5), 62-69.

Draper, M., Rowe, A., Douglass, S., Calhoun, G., Spriggs, S., Kingston, D., Humphrey, L., Patzek, M., Bartik, J., Lange, D., Reeder, J., Croft, B., Chen, J., Barnes, M., Trafton, G., Apker, T., Behymer, K., Ruff, H., Rothwell, C., Evans, D., & 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)). AFRL Technical Report (AFRL-RH-WP-TR-2018-0005).

Drexler, D. A., Takács, A., Nagy, T. D., & Haidegger, T. (2019). Handover Process of Autonomous Vehicles-technology and application challenges. Acta Polytechnica Hungarica, 16(9), 235-255.

Flach, J. M. (1988, October). Direct Manipulation, Direct Engagement, and Direct Perception: What's Directing What?. In Proceedings of the Human Factors Society Annual Meeting (Vol. 32, No. 19, pp. 1355-1358). Sage CA: Los Angeles, CA: SAGE Publications.

Flach, J. (2017). Supporting productive thinking: The semiotic context for Cognitive Systems Engineering (CSE). Applied ergonomics, 59, 612-624.

Franconeri, S. L., Alvarez, G. A., & Enns, J. T. (2007). How many locations can be selected at once? Journal of Experimental Psychology: Human Perception and Performance, 33(5), 1003.

Frank S.A. (2018) Time Delays. In: Control Theory Tutorial. SpringerBriefs in Applied Sciences and Technology. Springer, Cham. https://doi.org/10.1007/978-3-319-91707-8_13

Fern, L., & Shively, R. J. (2009). A comparison of varying levels of automation on the supervisory control of multiple UASs. Proceedings of AUVSI's Unmanned Systems North America 2009, 10-13.

Gibson, J.J. (1979). The ecological approach to visual perception. Boston, MA: Houghton Mifflin Co.

Gibson, J. J., & Carmichael, L. (1966). The senses considered as perceptual systems (Vol. 2, No. 1, pp. 44-73). Boston: Houghton Mifflin.

Gibson, J. J., & Crooks, L. E. (1938). A theoretical field-analysis of automobile-driving. The American Journal of Psychology, 51, 453–471. doi:10.2307/1416145

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Advances in psychology (Vol. 52, pp. 139-183). North-Holland.

Horrey, W., & Wickens, C. (2007). In-vehicle glance duration: distributions, tails, and model of crash risk. *Transportation Research Record: Journal of the Transportation Research Board*, (2018), 22-28.

Hutchins, A. R., Cummings, M. L., Draper, M., & Hughes, T. (2015, September). Representing autonomous systems' self-confidence through competency boundaries. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 59, No. 1, pp. 279-283). Sage CA: Los Angeles, CA: SAGE Publications.

Hutchins, E. L., Hollan, J. D., & Norman, D. A. (1985). Direct Manipulation Interfaces. Human-Computer Interaction, 1:4, 311-338.

Jagacinski, R. J. (1977). A qualitative look at feedback control theory as a style of describing behavior. Human Factors, 19(4), 331-347.

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C. M., Van Riemsdijk, M. B., & Sierhuis, M. (2014). Coactive design: Designing support for interdependence in joint activity. Journal of Human-Robot Interaction, 3(1), 43-69.

Klein, G., Wiggins, S., & Dominguez, C. O. (2010). Team sensemaking. Theoretical Issues in Ergonomics Science, 11(4), 304-320.

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

Larue, O., Juvina, I., Cox, M. T., Molineaux, M., Howard, B., Nichols, E., & Minnery, B. (2020). Coordination in homogeneous and heterogeneous teams.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human factors, 46(1), 50-80.

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. Nature neuroscience, 17(3), 347-356.

McCarthy, D. (2014). Choosing between centralized and distributed control system designs. Control Engineering Magazine and Newsletters. https://www.controleng.com/articles/choosing-between-centralized-and-distributed-control-system-designs/

McDermott, P., Dominguez, C., Kasdaglis, N., Ryan, M., Trhan, I., & Nelson, A. (2018). Human-machine teaming systems engineering guide. MITRE CORP BEDFORD MA BEDFORD United States.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological review, 63(2), 81. [PubMed] [Google Scholar]

Miller, C. (2014). Delegation and Transparency: Coordinating Interactions so Information Exchange is No Surprise. Virtual Augmented and Mixed Reality.

Designing and Developing Virtual and Augmented Environments Lecture Notes in Computer Science Volume 8525, 191-202.

Miller, C., Goldman, R., Funk, H., Wu, P., & Pate, B. (2004, June). A playbook approach to variable autonomy control: Application for control of multiple, heterogeneous unmanned air vehicles. In Proceedings of FORUM 60, the Annual Meeting of the American Helicopter Society (Vol. 3, pp. 2146-2157). Baltimore: American Helicopter Society International, Inc.

Mole, C. D., Lappi, O., Giles, O., Markkula, G., Mars, F., & Wilkie, R. M. (2019). Getting back into the loop: the perceptual-motor determinants of successful transitions out of automated driving. Human factors, 61(7), 1037-1065.

Norman, D. A. (1986). Cognitive engineering. User centered system design, 31, 61.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human factors, 39(2), 230-253.

Parasuraman, R., Molloy, R., & Singh, I. (1993). Performance consequences of automation-induced “complacency.” International Journal of Aviation Psychology, 3, 1–23.

Pawlak, W.S., Vicente, K.J. (1996). Inducing effective operator control through ecological interface design. International Journal of Human–Computer Studies 44 (5), 653–688.

Proffitt, D. R., Stefanucci, J., Banton, T., & Epstein, W. (2003). The role of effort in perceiving distance. Psychological Science, 14(2), 106-112.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE transactions on systems, man, and cybernetics, (3), 257-266.

Rasmussen, J., Pejtersen, A. M., & Goodstein, L. P. (1994). Cognitive systems engineering.

Rochlin, G. I. (1997). Trapped in the Net: The Unanticipated Consequences of Computerization. Princeton, NJ: Princeton University Press.

Rochlin, G. I., La Porte, T. R., & Roberts, K. H. (1998). The self-designing high-reliability organization. Naval War College Review, 51(3), 97–112.

Roscoe, A. H. (1984). Assessing pilot workload in flight: Flight test techniques (No. AGARD-CP373). Neuilly-sur-Seine, France: NATO Advisory Group for Aerospace Research and Development (AGARD).

Roth, E. M., Sushereba, C., Militello, L. G., Diiulio, J., & Ernst, K. (2019). Function allocation considerations in the era of human autonomy teaming. Journal of Cognitive Engineering and Decision Making, 13(4), 199-220.

Sage, A. P., & Cuppan, C. D. (2001). On the systems engineering and management of systems of systems and federations of systems. Information knowledge systems management, 2(4), 325-345.

Sarter, N.B., Woods, D.D., & Billings, C.E. 1997. Automation Surprises. Handbook of Human Factors & Ergonomics, Second Edition, Salvendy (Ed.), Wiley.

Shattuck, L. G., & Woods, D. D. (2000). Communication of intent in military command and control systems. In The Human in Command (pp. 279-291). Springer, Boston, MA.

Simon, H. A. (1996). The sciences of the artificial (3rd ed.). MIT press, Cambridge, MA. ISBN 9780262193740.

Stecz, W., & Gromada, K. (2020). UAV mission planning with SAR application. Sensors, 20(4), 1080.

Taylor, R. M. (1990). Situation awareness rating technique (SART): the development of a tool for aircrew systems design. In Situational Awareness in Aerospace Operations (Chapter 3). France: Neuilly sur-Seine, NATO-AGARD-CP-478.

Vicente, K.J., Rasmussen, J., 1992. Ecological interface design—theoretical foundations. IEEE Transactions on Systems Man and Cybernetics 22 (4), 589–606.

Vidulich, M. A., Wickens, C. D., Tsang, P. S., & Flach, J. M. (2010). Information processing in aviation. Human factors in aviation, 175-215.

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), Attention and performance VIII (pp. 239–257). Hillsdale, NJ: Erlbaum.

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

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2016). Engineering Psychology and Human Performance, Abingdon, NY.

Witt, J. K., & Proffitt, D. R. (2005). See the ball, hit the ball. PSYCHOLOGICAL SCIENCE-CAMBRIDGE-, 16(12), 937.

Witt, J. K., & Sugovic, M. (2010). Performance and ease influence perceived speed. Perception, 39(10), 1341-1353.

Woods, D. D. (1984). Visual momentum: a concept to improve the cognitive coupling of person and computer. International Journal of Man-Machine Studies, 21(3), 229-244.

Yakin, H. S. M., & Totu, A. (2014). The semiotic perspectives of Peirce and Saussure: A brief comparative study. Procedia-Social and Behavioral Sciences, 155, 4-8

APPENDIX A – Demographics Questionnaire

Simulation Evaluation Study Demographics Survey

Pilot #__________________ Grade / Rank: ___________ Active Duty? Yes ___ No ___

Service / Unit: ______________ Age: _______________ Gender: Male ___ Female ___

Current Position____

Handedness: Left ___ Right ___ Glasses? Yes___ No___ Hearing Aids? Yes___ No___

Highest Level of Education (Degree/Subject)

____

Aircraft Qualifications (Hours) (e.g., AH-64D, 850 hours, OH-58D, 120 hours)

____

Flight Hours

Total Flight Hours	Fixed-Wing Hours	Rotary-Wing Hours	Combat Hours	Night Vision (NVG) Hours	Night Vision System Hours	Weather Hours (IMC/IFR)	Instructor Hours	AMC Hours

Specific Command roles (e.g., Air Mission Commander, Air Weapons Lead)

____

Aviation specific qualifications (e.g., Instructor, Maintenance, XP, Safety, Tactics, WTI)

____

UAV & aviator experience (e.g., Ground Control Station, Air Mission Commander with organic UAVs):

Have you teamed with UAVs (e.g., Grey Eagle, Shadow, Reaper) during training and in combat?

Training – Yes _____ No _____ If Yes, approximate number of hours ________ hours Which UAV? _________

Combat – Yes _____ No _____ If Yes, approximate number of hours ________ hours Which UAV? _________

If you teamed with UAVs in training or combat, which Level of Interoperability (LOI) did you use?

LOI 2 _____ LOI 3 _____ LOI 4 _____ Not Applicable _____ (check all that apply)

Were you an AMC during missions where you teamed with a UAV? Yes _____ No _____

If Yes, how many hours were you an AMC when teaming with a UAV? _________ hours Not Applicable _____

Have you performed duties in a Ground Control Station as a payload or air vehicle operator? Yes _____ No _____

If Yes, describe your duties in the Ground Control Station and how many hours:

____

Number of months deployed to combat

____

What countries/areas of operation have you flown in (dates)?

____

Video game experience (consoles played, # of hours/week played, types of games)

____

Have you ever used VR headsets (e.g., Oculus Rift, HTC Vive)? Yes ___ No ___

If yes, briefly describe your experience (e.g., just a demonstration, play video games, used in training)

____

Civilian Aviation experience (e.g., CFII, Examiner, ATP, Aircraft ratings)

____

APPENDIX B – Post Vignette Questionnaire

Post-Vignette QUESTIONNAIRE

Date: ____________ Vignette # __________ Subject # _________

How confident are you that you secured the ingress route?

____ Not at all confident

____ Somewhat confident

____ Confident

____ Very confident

____ Extremely confident

How confident are you that you secured the LZ?

____ Not at all confident

____ Somewhat confident

____ Confident

____ Very confident

____ Extremely confident

What were the most difficult/challenging aspects of this vignette?

What mistakes might a novice/inexperienced AMC make in this vignette?

System/Automation Questions:

Which system capabilities do you feel contributed the most to your ability to perform this particular vignette successfully?

Were any important system capabilities missing or lacking that you would have used in this vignette?

Did you find any of the interface elements difficult or confusing to use? Did anything not behave in a way you expected? What important information or function is not available in the interface?

Did you find the vignette, and the way you were instructed to execute it, to be realistic?

To what extent did the manned and unmanned systems collaboration capability contribute to mission success?

_____ none

_____ very little

_____ somewhat

_____ helpful

_____ critical (could not complete mission without it)

APPENDIX C – Post Evaluation Questionnaire

(Post-Test Interview Questions Continued: p.1 of 5)

Date: ____________ Vignette # __________ Subject # _________

This is a semi-structured interview. Not all questions need to be asked. Additional questions/discussions may arise regarding specifics of the participant's performance, review of the data on the evening of day 2, and observations by team SME(s).

Announce Subject Number and Date (for audio recording)

Autonomy

- How confident were you that you knew and understood what was going on.
- Did the autonomous behavior of the UAVs ever surprise you
- Did you feel you could modify the behavior of a single UAV? Multiple UAVs?
    - If not, what would you have liked to modify?
- How difficult was it to modify UAV behavior?
- How much collaboration was there between the UAVs?
    - Did you think of them as a team or unit performing autonomously?
- How do you feel about the level of autonomy, too much, too little, just right?
- Did you feel you were a participant and a leader in the mission or just the leader?

HMI

- Please comment on the location and placement of controls (i.e., physical & touch)
    - What did you find difficult or disruptive?
    - What did you find effective and valuable?
    - What would you change?
- Please comment on the readability and understandability of information.
    - What did you find difficult or disruptive?
    - What did you find effective and valuable?
    - What would you change?
- After you had planned what you were going to do for each vignette, how much did you have to think about translating that plan into the 'language' or the commands of the system?
  ___ The language was the same as what I used in planning
  ___ There was some translation involved
  ___ I had to put some effort into putting it into 'system speak'
- When requested to make changes by the TOC, respond to a new threat, or track a target, did you have to stop and think/remember how to do it or were you just able to do it.

Date: ____________ Vignette # __________ Subject # _________

(Post-Test Interview Questions Continued: p.2 of 5)

Controlling multiple UAVs

- How would you compare performing a mission with 4 UAVs with performing a mission with 2 UAVs?
  - ___ Much more difficult
  - ___ More difficult
  - ___ The same
  - ___ Easier
  - ___ Much easier
  - Why?
- What were you able to do with 4 UAVs that you couldn't do with 2?
- Did you feel that you made the best use of the 4 UAVs?
  - If not, why?

Pilot workload.

- What mission execution tasks were the most difficult and highest workload?
  - What made them difficult?
  - What would you change to reduce the difficulty?

Situational awareness.

- Did you ever become confused or felt like you lost the mental picture of what each of the aircraft in the simulation were doing?
- Did the autonomous UAV behaviors make it difficult to understand and predict what is happening and what will happen next?

Trust.

- Once you gave a UAV (or a set of UAVs) a command, were you able to go do other tasks or did you constantly monitor what the UAV(s) was doing.
  - If monitoring, was it
    - ___ most of the time,
    - ___ half of the time, or
    - ___ just periodically until it completed its task.
- Did you agree with the plans that were generated by the system?
- Did you ever take direct control of the UAV because the plan did not make sense?
  - How significant was the change in workload when taking control?
    - ___ no change,
    - ___ moderate, or
    - ___ severe.
- Did you feel like you wanted to modify UAV trajectories?
  - If so, how would you have preferred doing it?

Date: ____________ Vignette # __________ Subject # _________

(Post-Test Interview Questions Continued: p.3 of 5)

- Was there any additional information that could have been presented that would have increased your trust/comfort in the system?
- Did you satisfactorily understand how the system's plan was generated?
    - Did it make sense to you?

<u>The Future</u>

Future systems

- In your opinion would a touch-screen work in a helicopter vibration environment?
- If touchscreens were available in the cockpit, along with traditional input methods (e.g., bezel buttons, cursor controller, etc.) would you use the touchscreens?
- Would adding system status information at the top of the sensor display (e.g., video feed) be desirable?
    - Rate 1 (undesirable) - 10 (very desirable).
    - What information would you like to see in this section? Examples: mission progress, fuel/weapon, etc.
- Would adding voice-controlled functions be desirable?
    - Rate 1 (undesirable) - 10 (very desirable).
    - How effective would it be in a noisy helicopter environment?
- Would adding gesture controls be desirable?
    - Rate 1 (undesirable) - 10 (very desirable).
    - How effective would it be in a helicopter vibration environment?
- Would adding eye gaze controls be desirable?
    - Rate 1 (undesirable) - 10 (very desirable).
    - How effective would it be in a helicopter vibration environment?

The Role of AMC in future systems

- How would you incorporate UAVs into air assault missions?
- What challenges do you envision for integrating UAVs into air assault missions?
- What challenges do you envision an AMC will have when managing multiple UAVs during air assault missions?
- Do you feel you had the capacity to keep track of and effectively employ/manage more than 4 UAVs? If yes how many more?
- How much communication and coordination with other units is required when performing your mission tasks?
    - Do you envision this being a significant challenge when managing multiple UAVs coordinating with other units?

Date: ____________ Vignette # __________ Subject # _________

(Post-Test Interview Questions Continued: p.4 of 5)

- Under what circumstances would you want to think in terms of controlling the UAV or the UAV's sensors/weapons as opposed to collecting information, identifying targets, tracking targets, and engaging targets?
- How would you feel about controlling your ownship in the same way that you control UAVs? I.e., just another asset that you happen to be occupying.

<u>Simulation Evaluation</u>

Training

- Was the training adequate for you to proficiently execute the first vignette (Vignette 1) with enough skill and confidence?
    - If not, what was lacking?
    - By the end of the evaluation, did you feel proficient in using the system and in accomplishing a mission
- Was the mission briefing detailed enough for you to understand what you were expected to accomplish during the mission?
    - What would you change?
- Was the level of classroom/table/viewgraph training too little, just right, or too much?
- Was the level of individual task training in the simulator too little, just right, or too much?
- Was the level of vignette training in the simulator too little, just right, or too much?
    - Would you have liked an additional (different) vignette to train on?

Vignettes

- Were the vignettes/scenarios realistic?
    - If not,
        - What was not realistic?
        - How annoying/disruptive was that unrealism?
        - What would have made it better? What would you change?
- Were the vignettes challenging?
- Was the terrain
    - Realistic
    - Large enough
- Was the behavior of the enemy assets/entities and non-combatant assets/entities realistic?
    - If not, did it distract/annoy you?

Date: ____________ Vignette # __________ Subject # _________

(Post-Test Interview Questions Continued: p.5 of 5)

- Do you have any recommendations for how we could improve the vignettes or make them more interesting?

Simulator

- Was the simulator comfortable? If not, why?
- Did the out-the-window scene add any value, distract/disorient you, or not really matter?
- Did you feel that the screens and hardware were located within adequate reach after you adjusted your seat?

Did the presence of the control station and others (including noise) in the room distract you?

Simulation Evaluation Experience

- How do you feel about the
  - The introduction presentation on the first day?
    - Did you feel that you understood the purpose of the evaluation and what was expected of you?
      - If not, what surprised you or what do you still feel unsure about?
  - The questionnaires and debriefs
    - Pre-evaluation (e.g., demographics, informed consent)
    - Post-vignette
    - Final debriefing
- How comfortable would you feel about recommending others to participate in the evaluation?
  - If you would be reluctant to recommend, why?

(End of Post-Test Interview Questions)