Employing Manned and Autonomous Machines: Implications for LVC-based Training Simulations

Tim Cooley, PhD
DynamX Consulting
Oxford, MA
TCooley@DynamX.com

Ivar Oswalt, PhD
The MIL Corporation
Lexington Park, MD
IOswalt@MilCorp.com

ABSTRACT

The idea and rudimentary instantiation of autonomous vehicles dates to Leonardo da Vinci circa 1500 (Wired Brand Lab, 2016). Da Vinci's self-propelled cart is the earliest driver-less vehicle and, while vastly inferior compared to today's autonomous vehicles, the end goal remains the same: to develop a vehicle that can navigate a complex world with no human intervention. While the 21st century is vastly more complex, 21st century technology is also enormously more capable. Unmanned Autonomous Vehicles (UASs) now function in many domains and are very desirable in warfighting operations due to the reduced expense and human danger. This begs the question of how best to utilize autonomous vehicles in the fight, what training approach should be deployed, and what metrics assessed? Understanding the ability of Manned-Unmanned Teaming (MUM-T) to operate in a coherent and efficacious manner to accomplish shared mission objectives requires well-defined, measurable metrics. Using an LVC training environment with automated (to include autonomous and artificial intelligence-based) agents (LVCA), this paper discusses metrics that are specific to each of the manned or unmanned systems, as well as others that provide insight into the man-machine interface and interaction (McLean et al., 2013). Understanding the MUM-T interactions is critical because unless the human confidence achieved in the autonomous platforms exceeds a definable threshold, the unmanned systems will be used sub-optimally. In addition, metrics that reflect the synchronization achieved in MUM-T must include the logical assignment of each asset-type to individual tasks and the proficiency of the combined team. Future research extends the LVCA training simulation concepts into other areas like analysis, test and evaluation, and experimentation.

ABOUT THE AUTHORS

Dr. Tim Cooley is President and Founder of DynamX Consulting, a veteran owned consulting firm. Dr. Cooley spent 17 years on the United States Air Force Academy faculty, both in uniform and as a civilian, holding numerous positions including the DMSO Modeling and Simulation Chair, Deputy Department Head, and Senior Researcher. As the principal investigator, he completed many innovative mathematical and cost analyses for the USMC, USAF, and OSD and has performed counter-terrorism research for Joint Staff. He has significant experience using Machine Learning and Neural Networks for pattern recognition/classification in the medical community and is currently researching standards development in Artificial Intelligence.

Dr. Ivar Oswalt is a senior modeling and simulation (M&S) analyst at the MIL Corporation who leads Plans, Policy, Standards, Education, and Outreach activities for the Navy's M&S Office along with providing M&S and VV&A support to the Naval Research Laboratory and others in the US Department of Defense M&S Community. He has developed M&S metrics and measurement approaches; M&S composability taxonomies; and defined and analyzed M&S requirements. Dr. Oswalt authored this paper in his personal capacity and the views expressed are his own and do not necessarily reflect those of The MIL Corporation.

Employing Manned and Autonomous Machines: Implications for LVC-based Training Simulations

Tim Cooley, PhD
DynamX Consulting
Oxford, MA
TCooley@DynamX.com

Ivar Oswalt, PhD
The MIL Corporation
Lexington Park, MD
IOswalt@MilCorp.com

INTRODUCTION

For hundreds of years mankind has chased the enviable goal of building machines that can intelligently operate without human intervention. Leonardo Di Vinci, circa 1500, designed a self-propelled cart that followed a predetermined path (Wired Brand Lab, 2016). This is thought to be the first autonomous vehicle design; however, the vehicle was most likely never built for perhaps many reasons.[1] However, the lure of the efficiencies, the continuous production without tiring, and the off-loading of monotonous tasks to machines continued to spur scientists to pursue the development of intelligent autonomous machines/vehicles.

Robots performing mundane repetitive tasks have been utilized in manufacturing for decades and artificial intelligence (AI) methods and practices have existed for decades. However, until relatively recently, the computing power, sophisticated algorithms, and data needed to drive the AI tasks was not available in a small enough form factor to fit on a reasonable sized vehicle. While the DoD/Industrial complex is not yet at the point of sending a mixed manned/unmanned-autonomous team into the fight, recent demonstrations (General Atomics, 2023) have highlighted the capabilities of autonomous vehicles, and that mixed team employment may be closer than we think.

In this paper we describe the criteria needed to understand the effectiveness of teams that include humans and semi or fully autonomous vehicles, working together collaboratively in real time, to achieve a common objective. After quickly discussing the demand signal for autonomous vehicles, assessing manned-unmanned team performance will be summarized, metrics developed, and then a notional concept of operations will be presented on how to use Live, Virtual, and Constructive (LVC) simulation-based training to improve the performance of these teams using the metrics developed. It works to leverage and extend previously conducted analysis and architecture development (McLean et al., 2013) by for instance, adding adaptive under the rubric of automated.

DEMAND SIGNAL FOR USING AUTONOMOUS VEHICLES

Certainly, there are benefits to autonomous vehicles (AV) in the battlespace. One of the most important is that sending an unmanned agent into the fight, eliminates the risk to a human operator. Less risk to well-trained human operators is always a preferred option. Secondly, AVs are considerably less expensive than their manned counterparts. While the final costs for AVs are yet to be determined, current military drones are approximately $30M at the highest (Atherton, 2023) while the most expensive fighter jet, the F-35, is approximately $80M per copy. (Tirpak, 2023). Therefore, 24 AVs can be purchased for the cost of 9 state-of-the-art fighter jets. While the savings may not be as great for ground or underwater vehicles, there will still be less costs for unmanned vehicles.

Another distinct advantage is that machines do not tire. This means that not only can a machine stay in the fight longer, but as long as all systems are operational and the vehicle has fuel, there should be no degradation in performance (and an intelligent vehicle could even perform better with more experience). A human will not only have both physical and mental performance degrade with persistent stimuli response but will have that same performance degrade faster when under significant stress such as the stress of combat. An AV is able to fight, and to fight at top performance for many times longer than a human. An AV can accept more situational data and process it more quickly than a human. It can

[1] A replica of the design was eventually constructed and is kept at the Clos Luce museum in France.

be argued that a human can better synthesize the situation and draw meaningful conclusions, however, it can also be argued that with advances in AI that advantage is becoming far less.

Therefore, there seems to be several benefits to immersing AVs into the battlefield whether it is on land, under the water, or in air and space and there are several ways one can contemplate man and AVs interacting. There is one scenario where automated machines are in a space interacting with humans that are in a different space or geographical area. An example is a number of AI empowered drones performing a search and rescue operation interacting with a control center in a central location. Another possible situation is when an automated agent acts in conjunction with a human in the same vehicle. An example of this is an AI system acting as the "back-seater" in a two person aircraft. This AI agent is an assistant to the pilot performing a host of functions to increase the aircraft's effectiveness. The scenario that we are concentrating on is that of AVs working with humans in a similar scenario (see Figure 1). The human becomes not only a combatant, but a battle manager as well piloting his own vehicle and providing input and information to the AVs as they battle. That is, the situations in which the human is immersed in the same theater and scenario as the AV/Automated agent.

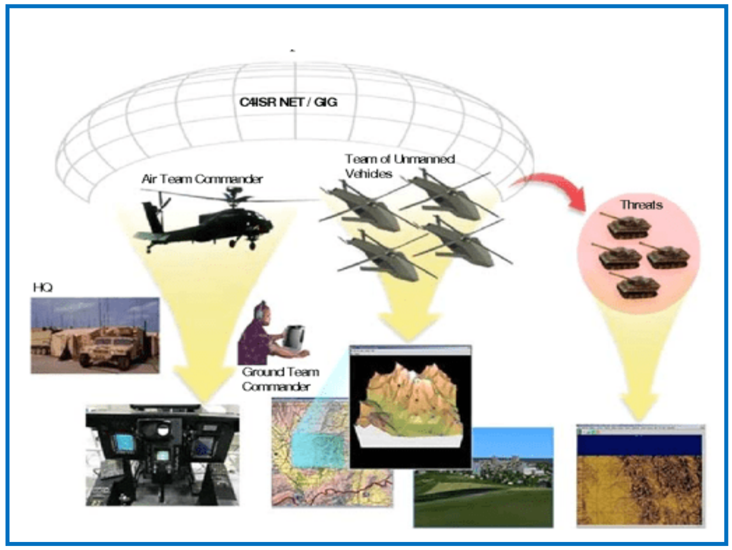

Figure 1 - Manned-Unmanned Team Example

THE ASSESSMENT OF MANNED-UNMANNED TEAMING (MUM-T)

Over the years there has been much study on measuring the effectiveness and performance of teams in many settings. While teams do many things, it is important to distinguish between "taskwork" and "teamwork". Gerald Goodwin, PhD, of the US Army Research Institute for the Behavioral and Social Sciences states that "taskwork" is the work teams must do to complete the mission or assignment while "teamwork" is the interrelated thoughts, feelings and behaviors of team members - comparable to the ABCs - that enable them to work effectively together (Goodwin et al., 2018). While completing tasks is required, what allows teams to function well - at a higher level than the sum of the individuals - is the teamwork aspect and the team cohesion required to function at that higher level. Goodwin says that how well people work together may be more important than how well they work on the tasks (Ibid, 2018).

As they work together, exchanging information and analyses, making predictions, formulating strategies, and executing decisions, there are several key dimensions that need to be used to measure the degree to which they are acting effectively as a team. Do they develop shared - and accurate - mental models of the situation, do they communicate to one another precisely and in a timely manner, do they leverage each other's strengths and avoid areas of weakness, do they trust each other to the point of assigning tasks without undo oversight, are they proficient, and finally do they win? All of these characteristics or outcomes can be measured, and training can be employed to improve them.

The aspect of cohesion has been measured in numerous ways in various scenarios and these efforts reported for several years (Beal et al., 2003; Chiocchio & Essiembre, 2009; Casey-Campbell & Martens, 2009). One of the key elements of team cohesion is the trust between team members. While humans can understand trust and know when we trust someone, describing it and measuring it is more problematic. When someone says, "I trust someone or something," it can mean different things to different people. The American Heritage Dictionary of the English Language, 5th Edition (2018) defines trust as: "Firm belief in the integrity, ability, or character of a person or thing; confidence or reliance." Attaining the belief in the ability and character of a teammate requires spending time with that team working tasks and learning mannerisms, abilities, communication methods, and pushing each other's limits until teaming becomes second nature. Goodwin also points out that team cognition - what teams think, how they think together, and how well synchronized their beliefs and perceptions are - is critical for teams to adapt to dynamic circumstances such as those found in the battlefield (Ibid, 2018). Team cognition is also built by the team working together, rehearsing scenarios repeatedly, being challenged with new scenarios and measuring how well the team performs, and exercising these scenarios until they can be accomplished almost without thinking. During this period and certainly after the trust has been built, being able to identify your teammate(s) (they are not imposters) is crucial to effective and proficient teaming. Devoting time to learn about a teammate and being able to identify the teammate are the basic building blocks to developing trust.

The military has the process of building teams well developed. Defined scenarios and exercises will take teams though many situations, each are executed until they are mastered, and then they are modified, and performed again. Certain team training is periodically exercised, and the teams are certified as proficient once they complete and pass the training. This has been the paradigm for decades and has served the military well.

So, is this process different for Manned-Unmanned Teaming (MUM-T, although MUMT is also used)? Is the process the same but the detailed execution different? If it is, and one might hypothesize that it is, how is it different and how do you build an effective MUM-T? What is the best way to employ LVC-based simulations to conduct such team building? Furthermore, and critically, how does one measure the effectiveness of MUM-Ts and is it different than a completely human team? These are some key questions regarding integrating AVs into the battlespace and using them effectively that we to start to address here.

Extending Standard Definitions and Bounding This Swim Lane

For this effort it is required to extend the standard definitions of LVC simulation. In the case of automated or adaptive (AI/ML-enabled) simulations, the key aspect that changes is the ability of both real people and software to provide inputs and make decisions. In the case of Live Simulation, for example, that means both types of systems consume and generate, as well as conduct, tactical option assessment and implementation.

So, we propose replacing “real people” with “operators” that can be real or synthetic. Thus, an updated set of definitions are:

- **Live Simulation** - A simulation involving operators controlling real systems within a simulated operational environment (e.g., a range).
- **Virtual Simulation** - A simulation involving operators controlling simulated systems within a simulated operational environment. (In the case in which the operators are synthetic/software, this is the same as Constructive Simulation.)
- **Constructive Simulation** - A simulation that involves operators controlling simulated systems within a simulated operational environment, but to which real people can provide inputs.

Table 1 - Example MUM-T LVC Types

Manned	Unmanned	Example
Virtual	Constructive	Soldier in a Virtual Combat Information Center (CIC) with a constructive representation of a supporting command.
Live	Constructive	A pilot flying an aircraft on a training range with an autonomous wingman being displayed on their Heads-Up Display (HUD) and as a track on their radar.
Live	Live	A scuba diver ingressing to a target with a set of autonomous Unmanned Underwater Vehicles (UUVs), all in formation, within an undersea range.

The focus of this paper, and two alternatives discussed in the training scenario described below, are Virtual/Constructive and Live/Live (see Table 1).

Understanding MUM-Ts and Training Metrics

Metrics need to be defined and measured for both the individual manned and unmanned systems, as well as for those that provide insight into the man-machine interactions and synergistic behavior (like the symbiosis that occurs between biological systems: e.g., Police Officers and their K9s). Understanding the man-machine link is critical because unless they can share a common understanding, communicate, know each other's strengths and weaknesses, and trust one and other, they will not be efficient nor effective. Ultimately, however, it's the proficiency of the team and the outcome that matters. The following sections discuss the development of metrics to measure how well an unmanned machine integrates with a manned machine.

Measures that Reflect MUM-T Characteristics

When it comes to the degree to which teams, in this case MUM-Ts, share the same or similar mental models, an approach developed within psychology provides a very useful construct (Van Rensburg, 2021). Fundamentally this framework has five primary areas of assessment, with four sub-areas within each. The primary areas are Equipment, Execution, Interaction, Composition, and Temporal (see Figure 2). It includes shared knowledge (e.g., what to do and how to do it, but also each other's knowledge, skills, and abilities), required interactions (e.g., information sharing and communications), and perceptions / temporal (e.g., deadlines and task urgency). While these can be easily tailored to the military context (i.e., equipment can easily be combat systems, capabilities, and gear), the level of their possession needs to be reflected by metrics that can be measured as they change during the training of a team.

Equipment
How to use other team member's equipment
What equipment is important for which tasks
The tools needed to complete our tasks
The technology needed to complete our tasks

Execution
Specific strategies for completing various tasks
How to deal with the task
How to best perform our tasks
The relationship between tasks

Interaction
How to communicate with each other
Sharing information with each other
How should we interact with each other
The best methods to communicate with each other

Composition
Each other's knowledge
Each other's abilities
Each other's skills for doing various team tasks
Each other's individual strengths and weaknesses

Temporal
Our deadlines
How quickly we need to work
Appropriately timing our work
Coordinating the timing of our work

Figure 2 - Five-Factor Perceived Shared Mental Model Categories

While this construct is comprehensive, two particular aspects merit further examination within the context of MUM-T training. The first is communication. As summarized well by Ming Hou (2021), "Communication speaks to how well teammates understand each other and how information is transferred in the team." Terms like clarity, directness, timeliness, means, can be translated into associated metrics. The next is that the strengths and weaknesses of the team are understood and acknowledged. Said another way, do the team members know each other's capabilities and is that knowledge accurate. This is especially important in MUM-Ts where these capabilities are likely to be widely divergent. For instance, a human may be able to consider a dozen alternative attack patterns from observing an enemy's attack, an AI/ML enhanced unmanned system might be able to consider hundreds. Such capabilities could be reflected (measured) via metrics on processing power, sensor performance, number of simultaneously considered options, etc.

Measures that Reflect Mutual Trust

One element of effective teaming is the concept of trust between and among the teammates. In April of 2022 Paul Nielsen proposed six dimensions of trust (Nielsen, 2022). While this is a useful model, a more behavioral approach that leads to easier assessment is presented here. Trust is manifest at three different levels and they all rely on each teammate processing information from another teammate without verification.

The first level of trust is that a teammate receives information and "notes" it or adds it to their database (digital or mental) for future use. For example, a driver in a semi-autonomous vehicle is presented a display (see figure 3) that shows another vehicle in the driver's blind spot. The driver notes that the vehicle is there and knows not to move into the other vehicle's path. The driver notes this without trying to look to verify the information.

The second level of trust is one where a teammate A receives information from teammate B and teammate A acts on the information, again without verify the validity of the information. For example, the driver above is backing the vehicle and an alarm tells the driver there is an obstacle in the way. The driver immediately stops and restarts once the path is clear.

The third level of trust is where teammates A receives information from teammate B that impacts both and teammate A allows teammate B to act without validating the information. In this case, using the semi-autonomous vehicle above, the vehicle alarms and states there is another vehicle coming from the left and the semi-autonomous vehicle automatically brakes or steers right to avoid a collision. While the driver of the vehicle can override the action when the alarm sounds, they choose to allow the vehicle to control and react (see Table 2).

Table 2 - Car MUM-T Example

Driver - Vehicle MUM-T Trust				
Level	**Scenario**	**Not Trusted**	**Trusted**	**Metrics**
One	Driver receives notice there is a vehicle close	Driver moves head / eyes to physically verify	Drive adjusts	Operator focus
Two	Driver receives alarm of obstacle in immediate path	Driver verifies by viewing camera or direct view	Driver takes immediate action	Operator focus and response time for action
Three	Vehicle alarms for imminent collision and takes corrective action	Driver overrides / attempts to override vehicle actions	Driver allows system to take action - no attempt to override	Operator focus and level of response

Using this model of trust, the metrics involve response time and area of focus of the personnel involved. The first level of trust listed above doesn't require a response. However, any response to the information provided which points to a verification attempt (shifting attention to validating the input) would be considered a lower level of trust than no response. While measuring this response would be highly situation dependent, using the example above with the semi-autonomous vehicle, one might use sensors to track eye or head movement to determine if the driver attempts to put "eyes on" the vehicle in the blind spot.

Figure 3 - View of Autonomous Automobile Control Center (Todayz News, 2019)

Since there is a response in level two trust, measuring this should involve reaction time as well as area of focus. Again, using the example given above in the description of the trust levels, metrics would involve the time it takes for the driver to take action based upon the vehicles input as well as measures of the driver's focus. Does the driver attempt to verify the information either by looking at a camera or moving head/eyes to physically see the object? Longer response time would point to less trust as would indicate an attempt to verify the information.

Measuring trust for level three above would take involves determining if one operator attempts to take over an automated response from the other. In the example above this would mean that the driver attempts to take over control of the vehicle as the automated agent is responding to the sensory input. Complete trust would be to allow the automated agent to take control ranging down to almost no trust if the driver automatically seizes control. This can be measured with sensors or cameras on the driver or a response from the automated agent that the other operator has taken control.

One of the critical aspects of MUM-T effectiveness is being able to verify that your teammate is the correct machine / unmanned system. A person has facial attributes that are recognizable, voice patterns that confirm who they are, physical characteristics, and shared historical experience that allow them to be identified. In the case of a machine, it is more problematic, yet ensuring the integrity (when a system performs its intended function in an unimpaired manner, free from unauthorized manipulation, whether intentional or accidental (NIST, n.d.)) of the components of the MUM-T is critical. Certainly, many systems have been attacked, compromised, and also coopted or replaced (thus enabling spoofing / the portrayal of a system they are not). Knowing the true identity of the unmanned system his would be critical at the beginning of entering battle. However, without a solid cybersecurity defense, a machine could be commandeered for nefarious means in the middle of battle and turn against its teammates. While, unfortunately, the protection and ramifications of system integrity or compromise are beyond the scope of this paper, it is important to acknowledge the critical role it plays, as the ideas presented here are further developed.

MUM-T Proficiency and Warfighting Outcomes

While having a shared mental model, communications, mutual understanding of strengths and weaknesses, trust, and integrity are all critical, even more so is the proficiency of the MUM-T and the outcome of the engagement. Becoming a proficient MUM-T will require the emphasis on some unique aspects of this particular type of team (like a greater requirement to recognize and effectively employ machine-unique traits (like excessive strength, computational speed, and enhanced sensory perception)). However, the fundamental features that define a capability of a human team remain a consistent benchmark for MUM-Ts. For instance, the indicators of proficiency of an infantry squad in taking an objective are the same for a MUM-T. Similarly, for the outcome metrics of winning / losing, success / failure.

Specific Metrics and Measurement Examples

Given the goal of effectively and efficiently employing manned and autonomous machines that have been trained using LVC-based Training Simulations, it is required to describe the characteristics of the type of autonomy being measured and to postulate associated metrics (see Table 3). It must be noted that the metrics listed below could be expanded significantly. For instance, communication metrics focus on correctness, completeness, and brevity, but there are many others that could be included (e.g., there are 19 listed in (Muszyńska, 2018)).

For the sake of these examples the unmanned component of the MUM-T is envisioned as tethered, remotely controlled, or teleoperated robot. An autonomous system would be one that is programmed prior to the mission, and which operates independently, but within a set of unchanging rules. An autonomous and adaptive unmanned system is one that dynamically changes its state of knowledge and behavior, symbiotically with the manned component of the team, using algorithms that implement artificial intelligence, machine learning, and/or emergent behavior.

Table 3 - Sample Metrics for the Unmanned Component of the MUM-T

<table>
<tr><th rowspan="2">Feature</th><th colspan="3">Type of Autonomy</th><th rowspan="2">Metric</th></tr>
<tr><th>Controlled</th><th>Autonomous</th><th>Autonomous + Adaptive</th></tr>
<tr><td>Shared Mental Model (MM)</td><td>It knows only what the human tells it.</td><td>It has a shared MM that is complete enough to act accordingly.</td><td>It incorporates and improves upon/enhances the accuracy of the MM shared by the team.</td><td>• % or +/- of the match between the true empirical state and the shared MM.</td></tr>
<tr><td>Communications</td><td>It does what it is instructed to do.</td><td>It follows its programing and executes the associated instructions accurately.</td><td>It understands what the human wants and incorporates and/or extends it (as needed) to achieve mission outcomes.</td><td>• % or +/- of the match between data provided and sensed.
• % or º difference between the amount of data needed and that provided.
• # of times data is repeated or superfluous data is provided.</td></tr>
<tr><td>Strengths and Weaknesses</td><td>The human in this team understands these characteristics.</td><td>It has a consistent/unchanging understanding of these characteristics.</td><td>It has an initial and evolving understanding of these characteristics which may change over time.</td><td>• % or º difference between characteristics possessed and their actual existence.
• % or º difference over time in the MUMT components understanding of these characteristics.</td></tr>
<tr><td>Integrity</td><td>Its software and systems are stable and can be securely validated.</td><td>Its software and system remains consistent and can be validated relative to established baselines.</td><td>Its software and system will adapt and change over time and so more complex validation tests will be required.</td><td>• % or º variance from known baselines when operationally deployed.
• Size, and changes in size of the potential attack surface</td></tr>
<tr><td>Trust</td><td>It does what I tell it to do.</td><td>It acts in ways that I expect.</td><td>It finds and takes the actions that I would have.</td><td>1. Accuracy - % or º provides accurate data or does what it should.
2. Reliability - % availability º defect free.
3. Resiliency - º to which adapts to input variability without error.
4. Objectivity - º to which system provides inputs/guidance that are unbiased.
5. Security - Like integrity, % or º variance from known baselines when operationally deployed.
6. Explain-ability - º to which data provided or actions suggested can be justified, after the fact.
7. Safety - º to which data provided or actions suggested do not cause harm.
8. Accountability - º to which data provided or actions suggested can be traced/associated with a particular entity or command.
9. Privacy - º to which data provided or actions suggested are shielded /restricted to authorized users. (1-9: Help Net Security, 2021; and Stanton & Jensen, 2021)
• Predictability - The confidence, prior to an action being taken, that the action taken by the unmanned component will be consistent with guidance previously provided. This includes dependability / reliability (Delahaye Paine, 2013).
• Relativity - The º or amount required given the scenario context (Lindley, 2023)</td></tr>
</table>

In the realm of trust, the predictability metric is particularly important. It includes the stability of the unmanned system in its provision of data, decisions, and actions that conform with the expectations/ intentions of the human operator within the MUM-T. When predictability is high, along with the other nine metrics of trust, then the employment of the MUM-T system can be optimized.

TYING IT ALL TOGETHER

The mission, well documented in the literature, is to employ a manned aircraft augmented with unmanned air vehicles (UAVs) to, as a team, ingress to a target, neutralize it, and then egress back to the rendezvous point (RAND, 2023; BAE Systems, 2023; UAV Navigation, 2023). In this example, five of the UAVs are configured for anti-air warfare (AAW) and five are configured to suppress enemy air defenses (SEAD). The team must work effectively together to employ their combined weapons' systems and abilities to the best effect against a state-of-the-art adaptive threat (see Figure 4). In addition, it is important to note that in this scenario, "the boundary between what a human does and what an autonomous system does during operation may shift during a mission." (Nielsen, 2022).

Figure 4 - Air Combat MUM-T Example

The team's first training mission is within a virtual / constructive simulation environment. While the outcome of this event was a success, blue force employment was sub-optimal. The UAV formations were not geographically distributed appropriately and in addition the aircraft's human pilot didn't trust the UAVs capabilities and so employed a kinetic kill option prematurely. So, additional virtual / constructive simulation-based training was conducted until the aircraft pilot had the requisite confidence in the accompanying UAVs. This training progression was then extended to a live / live simulation to exercise the MUM-T's capabilities in the most realistic environment. In this case, the scenario outcome was successful and used the combined strengths of the MMUT much more effectively.

Such sequenced and adaptive training programs are needed to, "equip airmen with the necessary skills to seamlessly integrate manned and unmanned assets, fostering operational synergy on the battlefield" (Harper, 2022). While contrived, this example outlines how MUM-Ts may improve their mutual understanding of capabilities and increase employment confidence using LVC-based training simulations. As the types of missions that employ the MUM-T construct expands to include, "reducing risks to aviators, extending the range of the fleet, enhancing combat capability, and serving as communication relay nodes, among other mission tasks" (Harper, 2022) the requirement for effective training will continue to grow.

FINDINGS AND CONCLUSIONS

The methods, metrics, and measures discussed and developed in this paper have not been fielded, and thus, the findings are limited. However, the need to assess MUM-Ts is as critical, perhaps even more critical, than assessing purely manned teams. As discussed here MUM-T assessment is different. Communication with a non-human entity is more difficult, there is no body language to read and the ability of the machine to understand nuances in questions is somewhat limited. Similar challenges present themselves in developing the same mental model between man and machine. Building trust can have an emotional connection and machines don't have emotions making the trust element of the team more difficult. Yet even though there are challenges, this paper demonstrates that meaningful measurable metrics can be developed, that they are different than metrics for purely human teams, and that the metrics and measuring these metrics is essential to understanding how well a MUM-T performs. Critical insights into the development and employment of LVC-based training within MUM-Ts using these metrics allows the adaptation of current simulation-based training approaches to now include autonomous agents in a meaningful way.

Finally, in conducting this work, it needs to be acknowledged that the best simulation training solution for MUM-Ts will vary by the phase of training and the blend of LVC being employed. Yet, it is our hope that the metrics can be defined and measured (or perhaps their values postulated or anticipated by subject matter experts a prior) across a canonical set of implementation options to construct an effective MUM-T initial concept of operations.

FUTURE WORK

The methods and ideas presented here are intentionally very general with limited specifics in order to appeal to a broad audience. The most obvious next step, then, is to operationalize the metrics and constructs discussed above in an actual LVC training environment. This requires specifying a training scenario or multiple coupled scenarios, applying the techniques given to develop the metrics, and then measuring and assessing the results. This assessment should include not only an assessment of the team immersed in the training but an assessment of the usefulness and applicability of the developed metrics. If necessary, the metrics would be tailored and the updated metrics applied to the next execution of that training scenario. Once that is complete, applying these constructs to a different space (if the first one was land vehicles, then the next experiment should be in the air or underwater space) is necessary to determine if the methods discussed here are broadly applicable. Finally, these concepts can be extended into other areas like analysis, test and evaluation, and experimentation. We look forward to reporting on these events.

REFERENCES

Atherton, K. D. (2023). Mass Market Military Drones have Changed the Way Wars are Fought, *MIT Technology Review.* https://www.technologyreview.com/2023/01/30/1067348/mass-market-military-drones-have-changed-the-way-wars-are-fought/

BAE Systems.(2023). *What is Manned-Unmanned Teaming?* https://www.baesystems.com/en-us/definition/what-is-manned-unmanned-teaming

Beal D. J., Cohen R. R., Burke M. J., McLendon C. L. (2003). Cohesion and performance in groups: A meta-analytic clarification of construct relations. *Journal of Applied Psychology*, *88*(6), 989-1004. https://doi.org/10.1037/0021-9010.88.6.989

Casey-Campbell M., Martens M. L. (2009). Sticking it all together: A critical assessment of the group cohesion–performance literature. *International Journal of Management Reviews*, *11*(2), 223-246. https://doi.org/10.1111/j.1468-2370.2008.00239

Chiocchio F., Essiembre H. (2009). Cohesion and performance: A meta-analytic review of disparities between project teams, production teams, and service teams. *Small Group Research*, *40*(4), 382-420. https://doi.org/10.1177/1046496409335103

Delahaye Paine, K.. (2013). *Guidelines for Measuring Trust in Organizations.* The Institute for Public Relations, Commission on PR Measurement and Evaluation; University of Florida. https://instituteforpr.org/guidelines-for-measuring-trust-in-organizations-2/

Editors of the American Heritage Dictionaries. (2018). Trust. *The American Heritage Dictionary Of The English Language,* 5th Edition. HarperCollins.

General Atomics Aeronautical. (2023). *GA-ASI Flies Multiple Missions Using Artificially Intelligent Pilots*. https://www.ga-asi.com/ga-asi-flies-multiple-missions-using-artificially-intelligent-pilots

Goodwin, G. F., Blacksmith, N., & Coats, M. R. (2018). The science of teams in the military: Contributions from over 60 years of research. *American Psychologist*, *73*(4), 322-333.

Harper, J. (2022). Navy tests manned-unmanned teaming technology with Super Hornets, drones. *FEDSCOOP*. https://fedscoop.com/navy-tests-manned-unmanned-teaming-technology-with-super-hornets-drones/

Help Net Security. (2021). *How do we decide whether or not to trust AI systems?* Retrieved from: https://www.helpnetsecurity.com/2021/05/26/trust-ai-systems/

Hou, M., Ho, G., & Dunwoody, D. (2021). IMPACTS: A Trust Model for Human-Autonomy Teaming. *Hum.-Intell. Syst. Integr.* 3, 79-97 .https://doi.org/10.1007/s42454-020-00023-x

Lindley J., Green D.P., McGarry G., Pilling F., Coulton P., & Crabtree, A. (2023). Towards A Master Narrative for Trust in Autonomous Systems: Trust as a Distributed Concern. *Journal of Responsible Technology*, *13*. https://doi.org/10.1016/j.jrt.2022.100057

McLean, T., Hoke, J., Vogl, T., & Schnell, T. (2012). LVCA: An Integrated Architecture of Live, Virtual, Constructive and Automated Elements for UAS Experimentation and Training. *Association for Unmanned Vehicle Systems International (AUVSI) Conference*, Washington D.C..

Muszyńska, K. (2018). A Concept for Measuring Effectiveness of Communication in Project Teams. *Journal of Economics & Management*, *33*, 63-79. doi:10.22367/jem.2018.33.04

National Institute of Standards and Technology. (n.d.) System Integrity. *Computer Security Resource Center Glossary*. Retrieved December 2023, from https://csrc.nist.gov/glossary/term/system_integrity

Nielsen, P. (2022). *Six Dimensions of Trust in Autonomous Systems.* Software Engineering Institute, Carnegie Mellon University. Retrieved from: https://insights.sei.cmu.edu/blog/six-dimensions-of-trust-in-autonomous-systems/

RAND. (2023). *Unmanned Aerial Vehicles*. https://www.rand.org/topics/unmanned-aerial-vehicles.html

Stanton, B. & Jensen, T. (2021). *Trust and Artificial Intelligence*. NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology. https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=931087

Todayz News. (2019) [graphic].*5 Things You May Not Know About Self-Driving Cars*. https://todayznews.com/5-things-you-may-not-know-about-self-driving-cars.html

Tirpak, J. (2023). $30 Billion F-35 Deal Will see Prices Rise, Deliveries Dip, *Air & Space Forces Magazine*, *106*(1). https://www.airandspaceforces.com/30-billion-f-35-deal-will-see-prices-rise-deliveries-dip/#:~:text=The%20unit%20cost%20of%20F,while%20annual%20deliveries%20will%20slip

UAV Navigation. (2023). *MUM-T: Manned-Unmanned Teaming*. https://www.uavnavigation.com/company/blog/mum-t-manned-unmanned-teaming

Van Rensburg, J. J., Santos. C.M., de Jong, S.B., & Uitdewilligen, S. (2022). The Five-Factor Perceived Shared Mental Model Scale: A Consolidation of Items Across the Contemporary Literature, *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.784200

Wired Brand Lab (2016). *A Brief History of Autonomous Vehicle Technology*. Ford Motor Company. https://www.wired.com/brandlab/2016/03/a-brief-history-of-autonomous-vehicle-technology/