ARL-TR-9431 ● Apr 2022

Final Report of DEVCOM ARL Director's Future Ventures (FY21): Novel AI Decision Aids for Decision Dynamics, Deception, and Game Theory (Summary Technical Report, Oct 2020–Sep 2021)

by Chou P Hung, J Zachary Hare, B Christopher Rinderspacher, Sue E Kase, Simon M Su, Walter Peregrim, Olena Tkachenko, Tomer Krayzman, Adrienne J Raglin, and John T Richardson

NOTICES

Disclaimers

ARL-TR-9431 ● APR 2022

Final Report of DEVCOM ARL Director's Future Ventures (FY21): Novel AI Decision Aids for Decision Dynamics, Deception, and Game Theory (Summary Technical Report, Oct 2020–Sep 2021)

Chou P Hung, J Zachary Hare, B Christopher Rinderspacher, Sue E Kase, Simon M Su, Adrienne J Raglin, and John T Richardson
DEVCOM Army Research Laboratory

Walter Peregrim and Tomer Krayzman
Oak Ridge Associated Universities

Olena Tkachenko
General Dynamics Information Technology

REPORT DOCUMENTATION PAGE						*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**						
1. REPORT DATE *(DD-MM-YYYY)* April 2022		**2. REPORT TYPE** Summary Technical Report				**3. DATES COVERED (From - To)** October 2020–September 2021
4. TITLE AND SUBTITLE Final Report of DEVCOM ARL Director's Future Ventures (FY21): Novel AI Decision Aids for Decision Dynamics, Deception, and Game Theory (Summary Technical Report, Oct 2020–Sep 2021)						**5a. CONTRACT NUMBER** W911NF2120-171, -173, -227
						5b. GRANT NUMBER R.0041203.80.1
						5c. PROGRAM ELEMENT NUMBER 622145BI2
6. AUTHOR(S) Chou P Hung, J Zachary Hare, B Christopher Rinderspacher, Sue E Kase, Simon M Su, Walter Peregrim, Olena Tkachenko, Tomer Krayzman, Adrienne J Raglin, and John T Richardson						**5d. PROJECT NUMBER** DFV20-HR-008
						5e. TASK NUMBER
						5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) DEVCOM Army Research Laboratory ATTN: FCDD-RLH-FC Aberdeen Proving Ground, MD 21005						**8. PERFORMING ORGANIZATION REPORT NUMBER** ARL-TR-9431
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)						**10. SPONSOR/MONITOR'S ACRONYM(S)**
						11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release: distribution unlimited.						
13. SUPPLEMENTARY NOTES ORCID IDs: Chou Hung, 0000-0003-1447-9828; J Zachary Hare, 0000-0002-3920-7442; B Christopher Rinderspacher, 0000-0003-1333-1653						
14. ABSTRACT This report describes efforts conducted in FY 2021 under the US Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory's (ARL) Director's Future Ventures project "Decision Dynamics, Deception, and Game Theory". To improve the effectiveness of decision aids for command and control of Multi-Domain Operations, it is necessary to develop artificial intelligence (AI) capable of assisting in complex decision-making. This project developed an AI test-bed, ARL Battlespace, for creating and investigating AI decision aids for complex reasoning. ARL Battlespace is a multiplayer network war game with teams of friendly and hostile human and AI agents. Initial results with hierarchical Bayesian modeling illustrate the potential for a framework for multidisciplinary development of AI with complex reasoning under scenarios with uncertainty, deception, and game theory. The project also began developing a framework for human–AI collaborative decision-making based on potential integration with the Battlespace Visualization and Interaction platform and with the Persistent Services Framework of high-performance computing. These results open research to improve the complex decision-making and collaborative capabilities of human–AI teams.						
15. SUBJECT TERMS STR, artificial intelligence, command and control, reinforcement learning, machine learning, decision-making, Humans in Complex Systems, Military Information Sciences, Network and Computational Sciences						
16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**		**19a. NAME OF RESPONSIBLE PERSON** Chou P Hung
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	UU	52		**19b. TELEPHONE NUMBER (Include area code)** 410-278-9255

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

1. Introduction

As part of the DOD's Artificial Intelligence (AI) strategy,[1,2] the US Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory (ARL) is developing research programs and technologies based on the Human Systems Adaptation strategy, including the goal of developing superhuman capabilities based on human–AI team decision-making and mutual adaptation. These new capabilities are necessary to address the Army's Multi-Domain Operations (MDO) strategy,[3] particularly its Penetrate and Dis-integrate phases during which AI-enabled decision aids can augment the commander's ability to tackle the high velocity and volume of information and the complex dynamics of the ground, sea, air, space, and cyber domains. A key challenge is that existing AI algorithms, including leading AI algorithms that are focused on specific problems in AI learning, are woefully inadequate for *complex* decision-making and have limited ability to generalize to MDO-relevant scenarios. Another challenge is that existing Army processes for doctrine and decision support do not integrate AI into the military decision-making process (MDMP),[4] and this gap is just beginning to be addressed by the Army's Automated Planning Framework (APF).[5] In addition, existing theories and technologies for human–AI team decision-making are limited to simple decisions, with very limited ability to provide AI transparency for complex decisions in depth, in which multiple dependencies, uncertainties, and information domains and actors intersect with complex human, materiel, and environmental dynamics. They also have limited ability to synergize with the tacit reasoning of human experts. Developing these capabilities requires an integrative and multidisciplinary research approach, including the development of AI test-beds for novel AI research and human–AI teaming.

For war-gaming, it is necessary to develop test-beds that can model decision-making across multiple echelons including tactical and strategic levels. Existing war-gaming decision tools such as Opsim,[6] AFSIM,[7] and OneSAF[8] can model and simulate many factors across multiple scales to predict outcomes based on strategies, materiel capabilities, and resources, but they suffer from the limitations of aging systems that can be difficult to learn for experienced Soldiers and that are not well suited for developing AI and human+AI teaming capabilities. The recent rapid rise in AI capabilities opens up research into the development and incorporation of novel AIs as decision aids for war-gaming. Recent improvements in AI reasoning (e.g., based on deep reinforcement learning) have been based on "open" games in which the state of the environment is perfectly known (e.g., checkers, chess, and go).[9] They are also based on limited cooperativity or deception. Even in cases with additional complexity such as environmental uncertainty (Angry Birds,[10] Atari[11]), there is limited decision complexity, flexibility, and transferability to multiplayer war-gaming (e.g., poker, Minecraft, Starcraft [Fig. 1]).[12,13] Although these models can explore decisions in depth, they are limited to conditions in which the potential values of choice outcomes can be easily measured and quantified. War-gaming environments pose a difficult and unaddressed challenge for AI learning because of the many sources of information uncertainty, not just from the environment but also from the human and AI agents. AIs need to adapt to changing rules and strategies, to rapidly mitigate unexpected hostile capabilities, and exploit new opportunities and friendly capabilities.[14] AIs also need to mutually adapt with their human teammates, and they need to have capabilities for tacit reasoning to synergize with human experts and for compensating for individual biases and heuristics[5] and changing cognitive states. Unlike classical approaches such as game theory, where the expected utilities of future states can be explicitly quantified for limited sets of actions depending on cooperative or noncooperative choices, war-gaming raises the possibility of interactions across environmental and social dynamics (including cooperativity and deception) and across multiple spatiotemporal scales and domains, which confound the AI's ability to learn how decisions tie to future state values.[15]

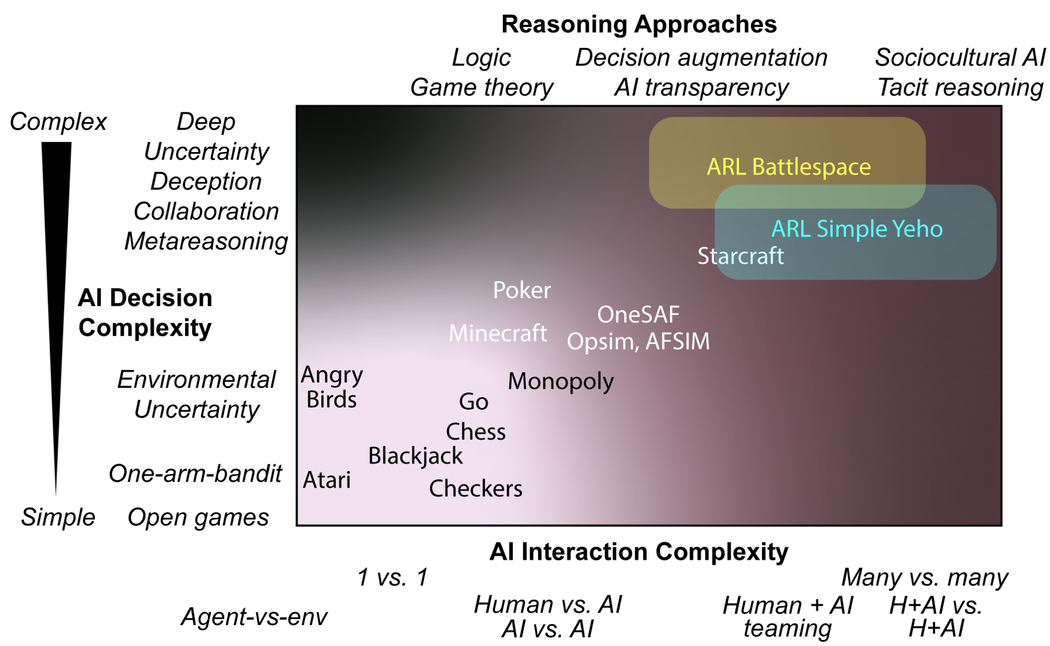

Fig. 1 ARL Battlespace within the broader AI research strategy

Addressing this gap requires a sustained foundational research effort with experiments focused on discovering principles and developing new algorithmic approaches for specific problems in decision-making, and the capability to tie these principles and algorithms back to MDO war-gaming. For example, in complex situations with imperfect knowledge and uncertainty,[16,17] an AI that provides a landscape of near-optimal solutions may be more helpful than one that provides a single "optimal" solution.[18] How this problem-solving ties to AI transparency also needs to be explored.[19,20] Experimentation of conditions such as near-optimality and uncertainty with new warfighter machine interfaces (WMIs) can lead to new algorithms, universal tools, and principles that better synergize the human+AI exploration of complex decisions.[21,22]

1.1 Army Relevance and Problem Domain

Part of the Army's Strategic Science and Technology (S&T) plan is to develop capabilities for "superhuman" decisions and actions.[23] For the Human–System Adaptation part of the S&T plan, the expected result is a partnership of uniquely human capabilities and the emerging capabilities of machines to maximize the speed and options to effectively respond to the complexity, intelligence, and dynamics projected in future sociotechnological environments of 2035 and beyond. It is expected that these research efforts will create new capabilities for human-guided machine adaptation, training of technology-savvy Soldiers, hybrid human-machine thinking, and next-generation human-systems integration and systems-level analysis capabilities. Because of the rapid ongoing changes to warfare, including constant technological change, achieving such capabilities requires developing a research program to advance AI and human-AI teaming specifically for complex decision-making.

As part of DEVCOM Army Research Laboratory's Director's Future Ventures (DFV) program, this project's goal was to develop an interdisciplinary program to address the gaps in the complexity of AI decision-making and in human–AI team decision-making. This included developing an AI research test-bed, ARL Battlespace, to abstract complex war-gaming decision-making to key elements so that AI and human–AI teaming development can focus specifically on the complex decision-making processes themselves, while avoiding the computational and conceptual limitations of physical realism and of present-day materiel and doctrine. This also included creating novel concepts for how to develop human–AI collaborative decision-making, to understand how to shape the information flow to enable mutual human–AI decision transparency, and to enable mutual adaptive learning under conditions in which both humans and AI have difficulty sifting through uncertainty and deception. Both explicit and implicit decision-making frameworks needed to be accessible via this abstract war-gaming test-bed so that AIs could learn and be challenged across multiple levels of reasoning. An appropriate level of abstraction was also needed to enable multiple types of research, including academic research at the intersection of neuroscience, AI, and decision theory, to advance the capabilities and complexity of AI decision-making and to improve its translation to military relevance.

1.2 Long-Term Goal

It is envisioned that in the Army of 2035 and beyond, command-and-control (C2) decisions are invigorated by decision aids that harness distributed AI capabilities across multiple echelons and that ingest data across all domains with complexity and speed that would overwhelm unaided Soldiers. The AI-enabled decision aids would enable forward-simulation and distributed training for the battlespace; enable adaptation and forward-projection of the possible effects of changes in conditions, friendly and hostile strategies, and capabilities during the Penetrate and Dis-Integrate phases of MDO; and enable after-action review of key decisions. The AI would provide transparency into its decision-making by enabling interactive visualization of both real and abstract decision spaces optimized for the individualized Soldier and situational context aligned to Army doctrines and shaping future doctrine. Conversely, the AI would co-adapt to the Soldier, learning how to navigate through complex decisions with insufficient, conflicting, or deceptive information, and reshape, refine, and present information for effective team decision-making. With AI agents as partners for the effective transformation and actionalization of data and leveraging of explicit and tacit knowledge, it is expected that distributed C2 commanders will be able to co-develop and coordinate courses of action across the many spatiotemporal scales and dimensions of MDO, and that the cross-domain interactions in tactics and strategy will be forward-simulated with increased resiliency to the dynamics of the environment, people, and strategies. In addition to the increased capabilities for complex decision-making, it is expected that the decision-making process itself will be accelerated by removing tedious calculations and other delays so that plans and strategies can adapt faster than real time to ever-changing battlefield and external (e.g., diplomatic, economic) factors.

To realize this future, the long-term goal of a program to develop novel AI for complex decision-making is to leverage continued advances across multiple disciplines. The development of "core AI" for reasoning, while rapidly progressing for simple decision-making, requires continued synergistic innovation and research from fields such as neuroscience and psychology to develop novel theories for reinforcement learning under conditions when the reward is difficult to assign to specific events or actions (e.g., because it was unclear to what degree of certainty who, what, when, where, or why to attribute the cause of the reward). Theories at the mechanistic level (e.g., how neuroglial networks may support tying disparate events to rewards) and at higher levels (e.g., how social rules can shape learning) are needed to bridge the gap between the present limited capabilities of core AI and the needs of C2 decision-making. Synergistic innovation and research are also needed to integrate the AI development with Soldiers' tacit reasoning processes to enable meta-learning and metareasoning of how decisions interact.

1.3 Objective of the DFV Project

The ARL DFV program is a mechanism designed to promote new directions in cross-disciplinary basic and applied research, address research gaps, and create new capabilities for the Army's mission. DEVCOM ARL fellows identified the Science of Analysis as one area of capabilities needed, with the potential for high payoff and the need for reimagining and expansion of existing programs and the need for new programs to establish new core competencies and to build up in-house expertise.

To create these capabilities, this DFV project had the primary objective of creating a new research program to develop novel AI for complex reasoning for C2 decision aids. This included developing an AI test-bed, ARL Battlespace,[24] to enable flexible development of novel AIs for complex reasoning specifically for MDO C2 decision-making. Existing war-gaming AI test-beds tend to be limited to simpler decisions, focusing more on tactical ground operations. For example, ongoing AI test-bed development efforts such as ARL Simple Yeho AI test-bed are focused on environmental realism, with multiple map layers including roads, foliage, and elevation to recommend decisions to a platoon commander for tasks such as route planning and Solder retasking. Because of the focus on the local terrain environment, the AI reasoning developed in that environment will be focused on fine-scale social and ecological dynamics, with sparser opportunities for in-depth training on collaborative and hostile decision dynamics. These problems of sparseness and complexity ("dinky, dirty, dynamic, and deceptive data"[25]) have confounded classic approaches to develop AI, especially for complex reasoning. Conversely, this DFV project's ARL Battlespace AI test-bed abstracts away the elements of local terrain to focus the AI learning and reasoning more specifically on *complex* MDO-relevant C2 reasoning in depth (multiple decision steps including more frequent opportunities for collaboration and deception). This enables more focused development of AI capabilities for complex multiagent (human, AI, and human+AI team) decision-making under C2 war-gaming contexts.

A second objective was to develop the conditions for effective human–AI teaming for complex decision-making by developing an effective WMI to research and develop how to present the AI's understanding and predictions and how to harness the human's understanding and predictions. This effort included leveraging and developing high-performance computing (HPC) resources for computational support, coupled with development of custom software for commercial 2D interfaces and mixed reality interfaces for decision-making (e.g., the Battlespace Visualization and Interaction (BVI) platform, based on the Augmented REality Sandtable [ARES] platform).[26] By developing multiple WMI approaches, we expected that these platforms would enable rapid prototyping research for complex decision-making as well as enable integration of our novel AIs with more established frameworks and teams for war-gaming training and simulation.

Together, we expected that these efforts in novel AI development, HPC computational support, and WMI development for realistic representations of decision spaces will create a new paradigm for development of human–AI teaming, paving the way for future advancement and modernization of multiple Army doctrines (MDMP, DOTMLPF,[27] METT-TC[28]) (Fig. 2).

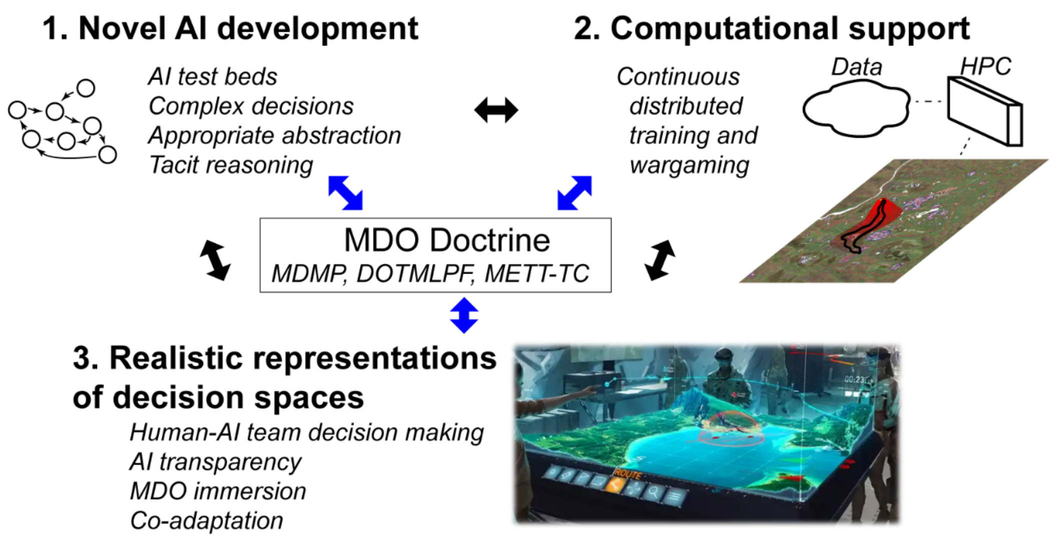

Fig. 2 Novel AI development within a broader human–agent team decision-making research strategy

This project resulted in the development of two research frameworks. First, it developed an AI test-bed, known as ARL Battlespace, for creating and investigating AIs for complex collaborative and hostile decision-making. Second, it recognized limitations in the current military decision-making process by conceptualizing a WMI for human–AI collaborative complex decision-making leveraging an Army and commercially developed battlespace visualization platform with potential connections to nontraditional HPC resources for enabling AI-enhanced war-gaming.

2. ARL Battlespace AI Test-Bed

Here, we describe our approach toward developing ARL Battlespace, an open-source flexible war-gaming platform that will lead to the development of novel reinforcement learning algorithms for decision aids.[29] Unlike other AI war-gaming approaches, our focus is on multiplayer decision-making. In particular, we are focusing on the gap in the theory and algorithmic capabilities for game theory with three or more cooperative and adversarial players. Whereas game theory concepts, such as Prisoner's Dilemma and Brinksmanship ("chicken"), are well developed for two players, they are not yet extended to three or more players, where the decision landscape can be complex due to saddle points and local minima that can confound efforts for reinforcement learning. Understanding and predicting Nash equilibria for three or more cooperative and adversarial players, under scenarios that are likely to emerge in warfare, require a flexible war-gaming platform that allows for the interdisciplinary exploration of such decision spaces. The war-gaming platform would also need to enable the development, understanding, and discovery of novel interactions and synergies between the players and the AI that enable the human to use the AI to quickly find optimal and near-optimal solutions. These solutions would enable the AI to learn from human patterns of decision-making and how to optimize its search of the decision space.

2.1 Framework

To enable these solutions, we developed a chess-like board game consisting of two teams, a red force and a blue force, where each team can have multiple coalitions (players) per team. The game is played on a common battlespace that is currently designed as a set of boards for each domain of the MDO. An example of the set of game boards is shown in Fig. 3, where we have considered an "Air" and a "Land" board. Each board is gridded into a set of cells, and the "Air" board is laid over the "Land" board to form a common battlespace. In this example, we have chosen to create square grids and only consider two domains. However, in general, the tessellation of the board can take any shape and be made arbitrarily small, while the number of boards is flexible to handle every domain in the MDO. For example, the "Air" board can consist of multiple boards that represent various levels of elevation. This formulation provides a general application programming interface (API) that allows for fundamental research advances in war-gaming since it can be customized to fit any war-gaming scenario.

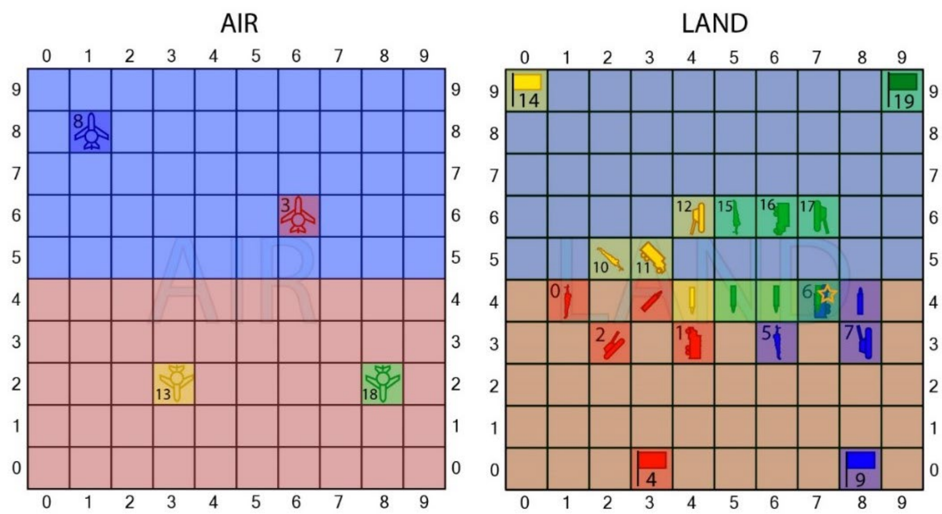

Fig. 3 ARL Battlespace AI test-bed for complex decision-making

Each coalition is assumed to have a set of pieces, which we call Units. Currently, we assume that there are four ground units and one air unit. The ground units consist of a Soldier, Tank, Truck, and a Flag, while the air unit is an Airplane. Each ground unit currently has the same capabilities (i.e., same set of actions and viewshed). However, the API is designed to enable customized capabilities for each unit of the coalition, making it easy to design specific scenarios.

The current rules and actions of the units are as follows. The Soldier, Tank, and Truck each have an orientation that describes their heading. Their actions consist of “doNothing”, “turnH”, “advance1”, “shoot”, and “ram”. “doNothing” implies that the unit stays in their location and does not change their orientation. “turnH” rotates the unit’s orientation by H degrees, where $H \in \{-135,-90,-45,45,90,135,180\}$. “advance1” moves the unit one cell forward in the direction of their orientation. “shoot” shoots a projectile in the direction of the unit’s orientation, where the projectile continues to advance by one cell forward until it either collides with another unit or travels outside of the game board. Finally, the “ram” action advances the unit one cell forward in the direction of its orientation, while attacking. The “ram” action is always advantageous as compared to the “advance1” action, since attacking can eliminate enemy units.

The Airplane unit has similar rules and actions as the Soldier, Tank, and Truck. These are “doNothing”, “turnH”, “advanceX,Y”, “shoot”, and “bomb”. The actions “doNothing”, “turnH”, and “shoot” are the same as the ground units. The action “advanceX,Y” allows the unit to move X cells along the East–West axis and Y cells along the North–South axis. The Airplane can also “ascend” and “descend” to take off and land. Finally, the “bomb” action shoots a projectile directly below the airplane onto the land game board. The Flag units are unable to move and are removed if they are captured.

The current implementation of the game play is simple. Initially, each coalition (player) places its units on their section of the game board. When there are multiple coalitions per team, the teams' portion of the game board is divided evenly between the coalitions. Note that the location of each unit is unknown to all the other coalitions. Then, each unit observes if there are any other units within its visible range, providing a fog-of-war scenario. We defined the observation range of each unit as one square from the unit's current location; however, the visible range can be customized according to the scenario and unit. Once each unit observes, the coalitions on the same team collaborate to identify the set of actions they would like to take for each of their units. This allows each coalition to observe their teammates' unit positions as well as communicate to coordinate their plans. Next, each coalition selects an action for each unit. Note that the actions chosen are only known to coalitions belonging to the same team. After the actions are chosen, a game resolution is applied that moves the units according to their selected actions and resolves whether any units have been attacked or collided with another. If a unit is attacked or collided with another unit, it is removed from the board. This process is repeated until the game is complete.

Completing the game depends on the underlying rules of the game, which are customizable to specific scenarios. Here, we studied two types of games: (1) Capture the Flag and (2) Annihilation. The goal of Capture the Flag is to maneuver the ground units into the enemy territory to capture the opposing team's flags, where the flag locations are unknown and must be discovered by exploration. The game is terminated once all the enemy flags are captured. The goal of Annihilation is to discover and attack all the enemy ground units. Here, the game is terminated once all the enemy ground units are discovered and eliminated. The underlying rules of each game are the same, but the best strategies for achieving each goal are different. In both types of games, there is high uncertainty due to limited visibility of enemy units and flags.

2.2 Pilot Experiment with Hierarchical Bayesian Modeling

Next, we report our initial results in developing an AI agent based on the idea of imitation learning using hierarchical Bayesian modeling constructed from human demonstrations. We start with a discussion of the data collection process, provide an analysis of the data, and finish with a heuristic approach that allows a simple AI agent to outperform a random agent.

2.2.1 Experimental Design

To learn human strategies, we had five human subjects combinatorially pair up together and play ARL Battlespace against two random agents for the two types of games discussed in Section 2.1 (i.e., Capture the Flag and Annihilation). During each turn, each random agent chooses an action for each unit i based on a fixed categorical distribution where the probability of taking an action $A_k^i \in \{a_1^i, \dots, a_{K^i}^i\}$ is $P(A_k^i) = \pi_k^i$ and K^i depends on the number of actions that unit i can take. Recall that the actions for each unit are described in Section 2.1.

Each game consisted of a pair of two human subjects versus two random agents, where at the beginning of each game, the human subjects collaboratively discussed their overall strategy for that game type. This led to the collection of 20 games, with 10 each for Capture the Flag and Annihilation, respectively. Once all the games were conducted, the game play data was analyzed to identify the human strategies.

2.2.2 Game Data Results and Analysis

The first approach to analyzing the game data was to study the frequency of actions taken by the human players. The frequency of actions is defined as

$$P(A_k^i|D) = \frac{N(A_k^i)}{T(D)},$$

where D represents the game data for either Capture the Flag or Annihilation, $N(A_k^i)$ is the number of times an action was taken by unit i over all the games, and $T(D)$is the total number of turns taken in all the games.

The frequency of actions is shown in Fig. 4 for the ground units (i.e., Soldier, Tank, and Truck) while Fig. 5 shows the action probabilities of the air unit (i.e., Airplane). The overall goal of the game dictates the action chosen, allowing us to determine the type of game being played. As seen in Fig. 4, the ground units of the Capture the Flag game are more likely to choose an advance and attack approach to search for the flag using the "ram" action. Additionally, an action of "doNothing" is also chosen more frequently. This is because once the flag is found by the team, the unit closest to the flag takes an action to purse it, while the remaining units do nothing. For the air unit, the human subjects were more likely to choose an "advance0,-2" action, which advances the unit into enemy territory to search for the flag.

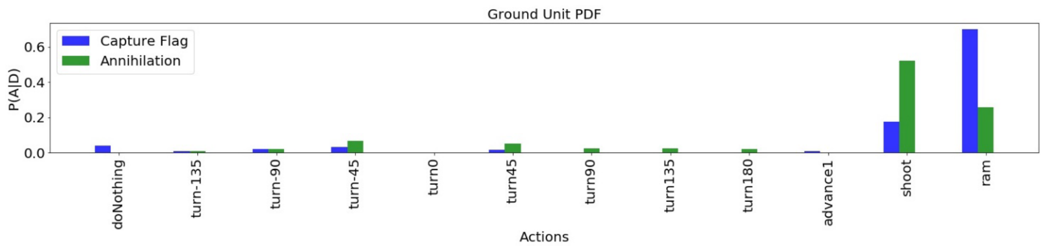

Fig. 4 The probability of an action conditioned on the type of game for all ground units generated from human game play

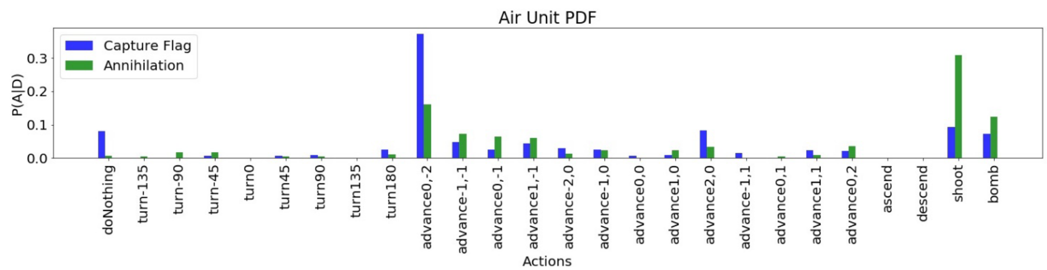

Fig. 5 The probability of an action conditioned on the type of game for air units generated from human game play

In the Annihilation game, the human agents were more likely to choose an attack action to eliminate the enemy targets (i.e., "shoot" for the ground units and "shoot" and "bomb" for the air unit). To further validate this strategy, the cumulative sum of the average number of projectiles per turn is shown in Fig. 6. Clearly, the Annihilation game results in a larger number of projectiles over the Capture the Flag games.

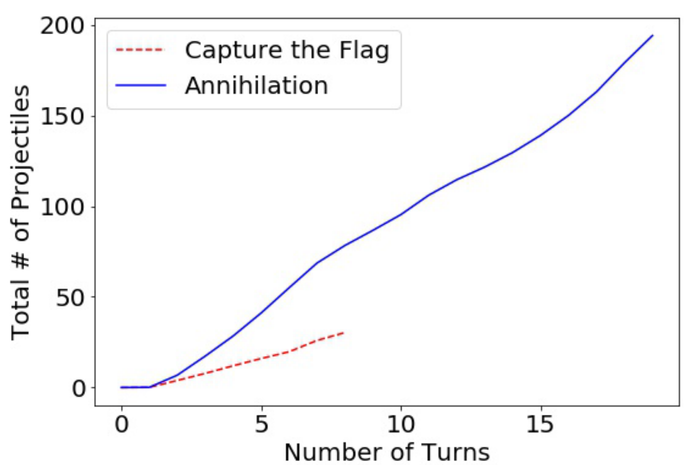

Fig. 6 The cumulative sum of the average of the total number of projectiles alive during each turn

Another difference between the two games is that the total number of turns for the Capture the Flag game is much less than for the Annihilation game. This is because the human agents are able to find the flags faster than they are able to find the enemy units and eliminate them.

Based on this simple understanding of how the human agents play the game against the random agents, we can follow a similar approach to learn the strategies to develop heuristics for a simple AI agent.

2.2.3 Performance of a Simple AI Agent Learning from Human Demonstrations

The algorithm for a simple AI agent is as follows. Initially, the agents randomly place their units in their designated areas of the board. Then, each agent identifies the state of each unit. Given the state and the goal of the game, the agent draws an action for each unit from a predefined probability distribution $P(A^i|S^i, D)$. This process is repeated during each turn until the game is over.

The predefined probability distribution follows a hierarchical Bayesian model. For ease of presentation, we have provided the theory in the Appendix. For the simplest case, we considered that the units could be in two possible states during each turn, $s_1 = Alive$ or $s_2 = Dead$. The probability distribution $P(A^i|S^i, D)$ is then defined according to Eq. A-1 in the Appendix and is similar to the frequency of actions presented in Figs. 4 and 5. We then implemented this distribution into two simple AI agents and played them against two random agents. As a baseline performance, we compared this against two random agents. In both cases, 1000 games were played, and the winning percentage was computed. By using the two-state probability distribution, the simple AI agents were able to win the game 84.5% of the time for the Capture the Flag games and 76.9% of the time for the Annihilation games.

Next, we considered a larger nine-state state space for each unit *i* defined as $S^i = \{(Fr0, E0, Fl0), (Fr1, E0, Fl0), (Fr0, E1, Fl0), (Fr0, E0, Fl1), (Fr1, E1, Fl0), (Fr1, E0, Fl1), (Fr0, E1, Fl1), (Fr1, E1, Fl1), Dead\}$, where $Fr0$ and $Fr1$ indicate if a friendly unit is observed by unit *i* or not, respectively; $E0$ and $E1$ represent if an enemy unit is observed by unit *i* or not, respectively; and $Fl0$ and $Fl1$ are whether the team has seen an enemy flag or not, respectively. Again, the probability distribution $P(A^i|S^i, D)$ is then defined according to Eq. A-1 in the Appendix and implemented into two simple AI agents. The winning proportion of the simple AI agents against two random agents was 89.4% for the Capture the Flag game and 82.3% for the Annihilation game.

A summary of the results is presented in Fig. 7. Interestingly, in both forms of the probability distribution $P(A^i | S^i, D)$ (i.e., the two-state distribution and the nine-state distribution), the Capture the Flag strategy outperforms the Annihilation strategy. This is because the agents in the Annihilation game are more likely to select the "shoot" action, which results in more friendly fires due to random initial placement. Therefore, it is more advantageous to take an attack-and-advance approach as a simple AI agent. Furthermore, the winning percentage increases as we considered additional states of the units. A possible direction for future work is to develop deep reinforcement learning strategies that will learn the definition and number of states needed to maximize the winning proportion, even against human agents, to provide suggestions for C2 in MDO.

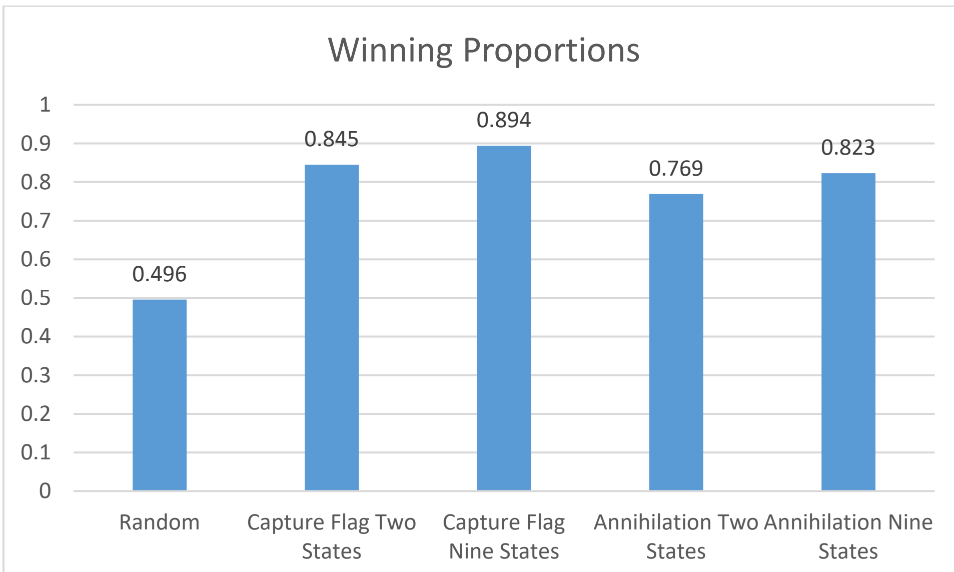

Fig. 7 Winning proportions of the simple AI agents

3. Example Scenarios with Complex Decision-Making

The key strength of the ARL Battlespace test-bed is its flexibility and adaptability to changing needs for MDO mission planning. Its abstract quality enables key decision processes and their interactions and dynamics to be compressed into a smaller gameboard and with more quantifiable human–AI interactions for development of AI for human–AI teaming. This enables the AI development to focus on the challenges of reward shaping for complex decision-making while reducing the impediments to learning that are due to nuisance factors (e.g., spatiotemporal scaling) which sparsify the decisions in time and space so that more effort (on the part of the AI as well as the AI developer) can be directed toward learning under uncertainty and deception at a variety of spatiotemporal scales. It also puts aside qualities of war-gaming interactions that may not be easily integrated into human–AI teaming (e.g., aspects of human psychology such as personal relationships) in favor of more tractable progress on AI reasoning development. In the following section we present a few example scenarios for challenging and developing AI for complex reasoning. These include examples with game theory, metareasoning, and cyber deception, touching on the variety of complex decisions not yet tackled or solved by existing AI algorithms. Because an AI-enabled C2 decision aid would be expected to *exceed* human-level decision-making, not just in speed but also in complexity, it is envisioned that such a C2 decision aid needs to be capable of solving most if not all these scenarios.

3.1 Breaching Scenario and Reimagining Game Theory

We begin by focusing on the gap between game theory and war-gaming in a simple breaching scenario, which is a classic problem in war-gaming that is often encountered (e.g., at bridge crossings, mine fields, and mountain passes [Fig. 8]). In the classic game theory concept of Brinksmanship ("chicken"), the friendly blue and green tanks are incentivized to cross the gap to reach the other side. Normally these tanks would coordinate their actions, but if the communication between the blue and green tanks is disrupted, the action of one unit (e.g., the blue tank) may lead to low payoff due to collision or friendly fire with another unit (the green tank). The scenario rapidly advances beyond classic game theory if it also includes elements of Prisoner's Dilemma, as it may be necessary for both the green and blue tanks to cross together to jointly attack the stronger red tank, requiring careful coordination. The presence of additional units (e.g., the green airplane providing observation, bombing, or jamming of hostile units such as the yellow Soldier providing possible reinforcement) enables further manipulation of dynamics and environmental constraints or opportunities on the decision-making. The airplane may also discover a second gap, or the "wall" may be permeable to create gaps (e.g., clearing the mines or establishing additional bridge crossings).

Behaviors learned at a coarse scale (e.g., 10×10 board) and context can be gradually generalized to finer scales and other contexts via reward shaping. Additional map layers can also be added for domains such as rapid underground transport to bypass walls in the ground layer. Environmental factors such as weather can also be included to alter maneuverability. Thus, even an apparently simple scenario can provide rich opportunities for manipulating factors that affect decision-dynamics and outcomes, and for exploring how interactions across different types of uncertainty can alter the decision landscape to create saddle points and local minima that can confound efforts at reinforcement learning. Understanding and predicting Nash equilibria for three or more cooperative and adversarial players, under scenarios that are likely to emerge in warfare, requires a flexible war-gaming platform that allows for the interdisciplinary exploration of such decision spaces. The war-gaming platform would also need to enable the development, understanding, and discovery of novel interactions and synergies between the players and the AI that enable the human to use the AI to quickly find optimal and near-optimal solutions. These solutions would enable the AI to learn from human patterns of decision-making and how to optimize its search of decision space.

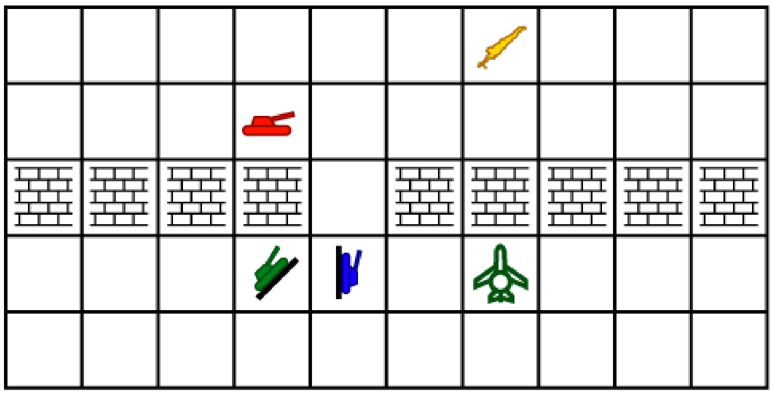

Fig. 8 Breaching scenario with enriched game theory conditions

3.2 Metareasoning Scenario, Mission Context, and Strategy

In the ARL Battlespace game, each player has a colored flag, and the game can be won by either annihilating all opposing ground units or by capturing all flags of the opposing team (a real-life equivalent is capturing all the key bridges or command centers). Depending on the state of the game, a commander may decide to alter the overall strategy (Annihilation vs. Capture the Flag) to achieve the win more quickly. For example, if one tank is already nearing one flag, it may be advantageous to redirect the remaining units to search elsewhere for the remaining flag (Fig. 9). Conversely, if a hostile unit is guarding the first flag, it may be better to prioritize capturing that flag so that the search for the second flag can be more efficient. This unarticulated reasoning, or "tacit reasoning", is often engrained in naturalistic human decision-making, and it is an AI capability that needs to be developed so that AI can participate effectively in human–AI team decision-making and so that AI development can begin to have tools to aspire for the creativity of human decision-making.

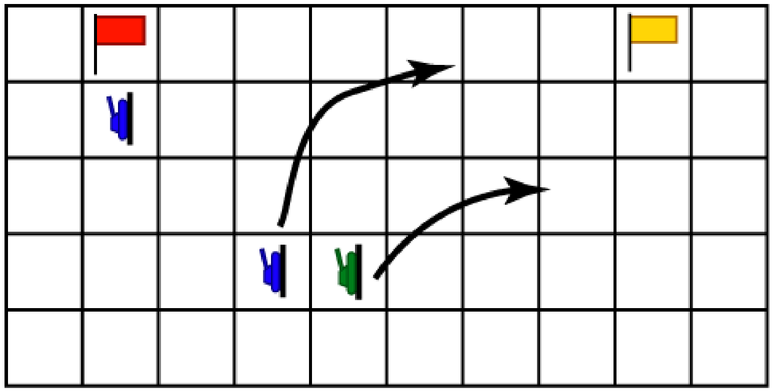

Fig. 9 Metareasoning flag scenario with tacit reasoning and task reallocation

For AI development, this would require that an additional higher-level reasoning agent be constantly monitoring the state of the game to make choices for switching strategies and to communicate this to the agents controlling the individual units. Metareasoning includes processes to monitor the steps involved in reasoning and to balance the criteria that factor into the outcome of the activity. In addition, metareasoning incorporates the uncertainties of different information to produce more meaningful and contextually appropriate decision recommendations. Incorporating metareasoning can allow constraints and various decision-making approaches to be weighed to provide different options for courses of action. For example, alternative metareasoning-based choices could decide whether to prioritize exploration versus attacking known hostile units versus defense, which maneuver strategy to deploy, or how to reallocate tasks given the observable positions of hostile forces. Because of the small grid size of the ARL Battlespace environment, the games can be played quickly, resulting in frequent opportunities for metareasoning to be used and opportunities for AI to learn to combine and predict interactions across multiple types of metareasoning approaches. Because the abstract environment increases the frequency of opportunities for the AI to learn how strategies interact, this would enable AIs to learn higher-order strategies such as the need to balance interactions across strategies, capabilities, and task requirements, to maintain freedom of choice and to produce strategic ambiguity to confound the opposition. Overall, the benefit of this approach is the improvements to decisions by adding the control and monitoring mechanisms that come with including a metareasoning agent that balances the actions and the environmental constraints.

3.3 Simple Deception and AI Theory of Mind

A key aspect of adversarial decision-making, particularly in warfare, is deception. Deception can occur across multiple levels including strategy, observable information, and unit capabilities and locations. The limited observability of units in ARL Battlespace naturally creates opportunities for deception, and the capability of airplanes to explore deep in hostile space provides opportunities to uncover deception about unit positions. Figure 10 illustrates an example of a simple deception scenario in which the friendly blue and green units attempt to cross to the other side. The friendly Soldiers at the lower left begin by firing missiles through the left gap because their agents reason (via AI Theory of Mind of the opposing agents[30]) that, upon seeing the missiles, the hostile agents will infer that the friendly forces are preparing to attack through that gap. This deception, by focusing the hostile agent's attention and planning to the left gap, biases them away from the right gap and creates an opportunity for the blue and green tank to enter from the right. By designing the scenario with two gaps, the scenario builds upon the two-alternative-forced-choice tasks of classic psychology, enabling the application of sensitive psychological tools for decision analysis and the development of animal models for neurophysiological and behavioral dissection of the underlying cellular and molecular mechanisms that govern context-dependent learning and decision-making[31] for deception. For example, one could introduce factors to bias the friendly or hostile decision-making (e.g., by manipulating the noisiness of sensors or by manipulating commands from headquarters), or apply methods such as optogenetics and chemogenetic tools to understand how the neural representation of others' perceptions, beliefs, or strategies (e.g., in the anterior cingulate and orbitofrontal cortex) contribute to decision-making computations (in the prefrontal cortex).[32,33] Such investigation could also uncover factors that determine single-mindedness, heuristics, and implicit bias versus openess to alternative hypotheses, which could help determine how best to reallocate tasks under specific conditions (e.g., when an individual is biased toward hierarchical command structure, he may be less open to pursuing sensor evidence that contradicts commands from headquarters). Such inherent biases, heuristics, and tacit reasoning are a natural component of human reasoning[34] and are anticipated in our interactions with others; it may be beneficial for AI theory of mind to include such bias compensation and expectations to optimize human+AI teaming.

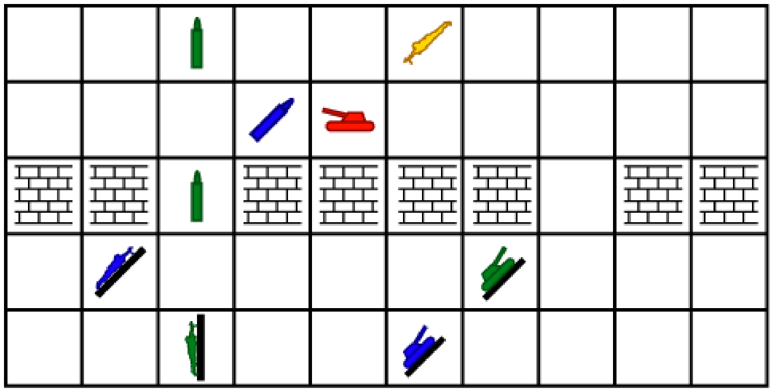

Fig. 10 Simple deception scenario requiring AI Theory of Mind

3.4 Cyber Deception, Multi-Domain Integration, and Believability

In human decision-making, information from different domains can combine to produce unexpected effects. The psychological McGurk effect[35] is when a strong temporal synchrony between the mouth gesture "ga" and the auditory syllable "ba" combine to produce the illusory percept "da". Although multisensory integration does not appear to have been explored in C2 decision-making, the confluence across multiple domains in MDO, particularly its high volume and velocity in the Penetrate and Dis-integrate phases, may produce unexpected nonlinear cross-domain interactions (this may contribute to the "fog of war"). Figure 11 illustrates an example in which a combination of actual evidence (missiles) and tank decoys (resulting from a man-in-the-middle [MITM] cyber attack) could synergize to compel the hostile units toward the left gap. It is a general strategy to create converging lines of evidence for cyber deception, yet specific patterns of deception may be more effective than others. For example, the brain is thought to group similar or related evidence into chunks for efficient processing (e.g., Gestalt grouping) so that it can overcome information bottlenecks (e.g., process more than seven nominal items, thereby reducing the impact of individual items). If carrying out each instance of cyber attack incurred a certain cost or risk, it may be beneficial to understand how to distribute these costs across cue signatures to deliver the most effective impact with minimal risk (e.g., the MITM attack would probably be less effective, or even counteractive, if it produced missile decoys). It may also be informative to understand how different combinations of cues may be differentially perceived by different Soldiers. Commanders with different biases or at different roles or echelons may perceive, interpret, or act differently on the same combination of evidence (e.g., a decoy's effectiveness is likely to depend on its distance to a target commander and relevance to his decision process). More advanced strategies may include active defense (e.g., via a "honeypot" strategy

[Fig. 12]) to improve the effectiveness of the cyber deception.[36] To deliver superhuman capabilities for MDO, an AI decision aid may need to assist in generating believable decoys across multiple domains based on the instantaneously available evidence, rapidly adapt these presentations at the speed of cyber networks, and maintain coherence between the virtual and real worlds in order to maintain the effectiveness of the illusion.

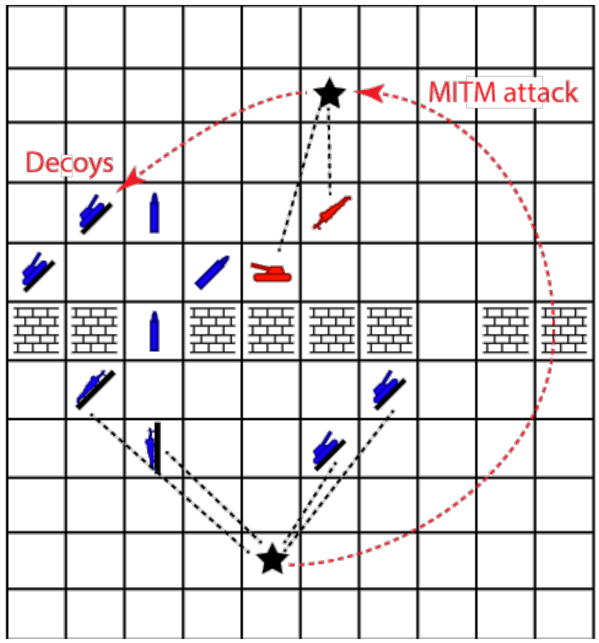

Fig. 11 Cyber scenario with man-in-the-middle attack

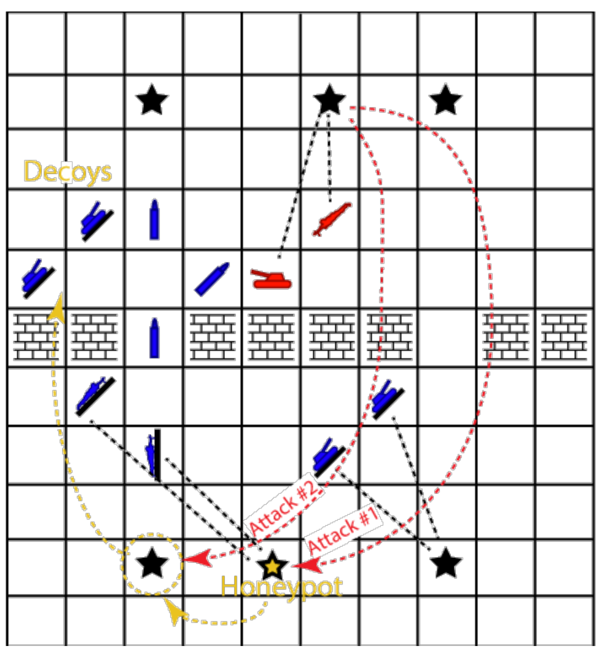

Fig. 12 Cyber scenario with honeypot

4. Human-AI Collaborative Complex Decision-Making

The ARL Battlespace AI test-bed described in the previous section offers the flexibility needed for AI development and testing by abstracting the battlespace terrain into a grid-like environment without realistic representation. For example, Fig. 8 shows a wall-like obstacle represented as several grid blocks associated with an environmental constraint condition applied during unit interaction. Both human teams and AIs play the game within the common bi-level gridded battlespace. The human players interact with ARL Battlespace by inputting text-based coded commands in a console window. This command line interaction and display accelerates the AI algorithm development process and sets up potential connections to computing resources for large-scale real-time calculations required for AI-enabled war-gaming. Conceptualizing a user interface for an AI war-gaming test-bed, such as ARL Battlespace, and establishing pipelines to external computing services constitute the foundational components of the second objective of the DFV—to develop a WMI for complex decision-making.

A model of the military decision-making process across echelons and operational levels forms the basis for developing an effective WMI for human and AI war-gaming. In traditional war-gaming, commanders utilize a common map-based operational terrain and model how combinations of factors within the MDMP produce courses of action (COAs), possible counter-actions, resource usage estimates, and predicted outcomes.[4] Over days or weeks, the MDMP process leads to a refined set of COAs that make certain assumptions about the operating environment, including terrain, weather, and the availability and capabilities of units in setting the theater (i.e., shaping activity in support of major combat operations).

Although MDMP assists command staff in understanding an operational environment and considering an operational approach, the process has many limitations such as time intensiveness, rigidity of the assumptions, limited opportunities for training across scenario variations, and few opportunities for integrating AI guidance into the decision-making process. Traditionally, the success of a mission is directly related to the ability of command to execute the MDMP. However, given the increased complexity of MDO with its vast array of mission command systems and processes, integration and synchronization of all activities associated with operations are becoming increasingly difficult to the point of humanly impossible. The lack of planning expertise resulting from a deficient MDMP can lead to desynchronized and dischordant operations and ultimately cost the lives of Soldiers.

The ability to visualize the battlespace is not specifically described in the MDMP, yet it obviously plays an important role in the decision-making process. Recently, new systems and technologies integrating advanced visualization capabilities have been developed that improve situational awareness and therefore enhance decision-making processes. Army examples include Nett Warrior,[37] which enables dismounted warriors to visualize nearby friendly and hostile forces while collaboratively planning tactical missions based on the local terrain. Although this technology extends the radio and digital mapping to the dismounted warrior, it lacks an underlying AI engine to provide decision assistance. BVI is another example of Army technology that enables distributed collaboration for mission planning with both 2D and 3D visualization capabilities of a common operating picture from arbitrary viewpoints and a wide selection of devices.[38] The BVI architecture can be formulated to pull in external computing services such as analytic pipelines, models, and AI engines.

Currently, MDMP does not incorporate AI guidance into the overall mission planning approach. The Army's APF[5] begins to address AI-assistive decision-making by inserting autonomous technologies into the MDMP workflow. Command staff can receive contextual assistance during mission planning and COA development through APF's digital plan representation, plan creator, and plan monitor tools. Mission execution and estimation capabilities provide automated assistance for improved decision tracking and support activities by monitoring planned versus actual progress of the mission. Although APF introduces a foundational level of automation into the MDMP, it lacks the advanced visualization and user interaction capabilities offered by Nett Warrior and BVI.

Aside from MDMP, recent efforts to integrate AI into the decision-making process have included a number of approaches,[12] with some success in modeling the human decision-making process. In general, AI has had some success for problems with limited decision variables, such as resource allocation,[6] flight simulators,[39] and simpler scenarios. Ongoing challenges include the need to improve the capability of AI to tackle complex decisions with multiple actors, incomplete and possibly conflicting or deceptive information, changing unit action and environmental properties, and the need to visualize the consequences of these decisions across many spatial and temporal scales.

4.1 Required Advancements for Future MDMP

MDMP limitations to support complex decision-making for MDO highlight the need for improvement in three areas. First, there is a need to integrate AI-generated guidance and assistive decision-making support into the MDMP. This includes both further development and integration of AI into battlespace decision planning, as well as further improvements in the explainability and transparency of the AI's decision-making process.[20] Second, there is a need to integrate the decision analytics with the power of HPC at the strategic level as well as the tactical edge when possible. This would enable leveraging the power of an HPC system to improve modeling, analytics, and computation time, while integrating and synchronizing information from across all theater domains. Finally, there is a need to develop more accurate and interactive representations of the decision space using advanced visualization technologies such as mixed reality. Rather than simply displaying a 2D rendering of the terrain at a fixed timescale, there is a need to visualize how decisions across different domains interact and leverage mixed reality to both improve the throughput and depth of the understanding and enable insights not possible with flat displays.

The MDMP lays at the core of the Army's design methodology for applying critical and creative thinking to understand, visualize, and describe problems and approaches for solving them. As the proven analytical process for problem solving, limitations of the MDMP as described previously must be overcome in order to quickly develop a flexible, tactically sound, and fully integrated and synchronized plan that increases the likelihood of mission success with the fewest casualties. The following subsections describe potential improvements to the MDMP to support human–AI collaborative decision-making.

4.1.1 AI-Directed Decisional Guidance

Novel AI-enabled WMIs are needed to both leverage ongoing advances in AI decision-making and contribute to AI learning for complex adaptive decision-making. The development of AI decision aids will provide increasingly capable suggestions of possible COAs by pooling information across all domains and computing risks and expected rewards for human and AI agents. There are several limitations of existing AI, particularly for complex and adaptive decision-making with uncertainty, with collaborative and adversarial human and AI agents. Modeling multiagent collaborative and adversarial decision-making can be particularly complex because of its recursive nature in which other agents are part of the model,[40] requiring dynamic and evolving estimates of decision features, individualized values, risk aversion, memory, and attention. These situations of high uncertainty, complexity, and dynamics are areas where humans excel and where appropriately designed interfaces for human–machine teaming can provide accelerated and more effective decisions. For effective teaming, the novel WMI should help the Warfighter to sift through complex information and help the AI to discover implicit rules for decision-making. Here we provide case examples of how human–machine teaming can be effective.

Complex decision-making as needed in multi-domain war-gaming is an immediate challenge for developing effective AI decision aids. The success of recent AIs in games such as go and chess are based on games with complete knowledge of the existing state of the world (i.e., "open" games), whereas war-gaming typically includes incomplete (e.g., Starcraft), uncertain, and/or deceptive information about the operational environment. The lack of knowledge makes it difficult for AI agents to calculate the risk-reward profiles of future actions due to the uncertainty in the state of the world, state of the different actors, and the effects of the actions taken.[41] Uncertainty also limits the ability of an AI to estimate the risk-reward profiles of the other actors, which are needed to calculate effective game theoretic strategies. It is not uncommon for AI to be overwhelmed by the breadth of possible optimal and near-optimal choices[18] (i.e., selecting the wrong choice due to limited information), since humans employ heuristics to make efficient choices and prediction when developing strategies for effective exploration of hidden information.[34] To assist the development of the AI's capability for implicit knowledge and exploration, novel WMIs need to explain and present the decision landscape effectively to allow the Warfighter to quickly and naturalistically navigate through possible choices, while enabling the AI to opportunistically learn from human decision-making without imposing cognitive burden.[42]

Another fundamental challenge for developing AI-enabled WMIs is how to effectively integrate and display information across all five domains in MDO, particularly space and cyber, as information across these domains has disparate spatiotemporal scales.[43] For cyber, the scale and speed of the decision-making can be faster than human capabilities to process and understand, requiring human input to guide semi-automated decision-making and an AI that implements strategies for offensive and defensive deception. The WMI needs to be able to display the decision landscape in such a way that a small list of optimal and near-optimal decision strategies are explainable (e.g., via a decision tree). This should include estimates of the future states and risk-reward profiles of key agents under uncertainty[16] to allow effective game theoretic decision-making to be co-developed and mutually understood.

These challenges inform the possible design of effective WMIs. Namely, we need the capability to ingest information from disparate sources (including from other nations' decision aids) and an architecture that can host the computational power to integrate this information, while also handling the underlying AI computations (both for learning and for deployment). We also need to co-develop an interface and algorithm design that opportunistically harnesses the strengths and mitigates the limitations of human and AI agents.

4.1.2 Computationally Informed Decision-Making

Substantial computation power is needed to process and record all components, entities, and state spaces during complex decision-making in MDO war-gaming. Past, present, and predictive modeling from accumulated data sets of dynamic state spaces requires leveraging HPC resources for generating analytic insights and creating representations useful in complex decision-making contexts.

One approach for implementing an HPC analytic workflow uses Persistence Services Framework (PSF). PSF is a recently available distributed virtualization solution that enables nontraditional access to high-performance computing services through a web-based front end, unlike traditional HPC environments where computational nodes are allocated to users in batch mode for a specific period of time. Additionally, PSF can provide distributed and continuous access to data, databases, containerized toolsets, and other hosted platforms.[38]

In an example PSF approach, a simulation engine connects to PSF for recording all decisions made by both the humans and AIs. This allows analysis of the decision-making behavior occurring during mission planning and COA development, as well as identification of decision-making patterns and strategies for developing competitive and realistic war-gaming scenarios. A battlespace visualization platform can be hosted on PSF and uses a messaging protocol to update all connected device interfaces. State information from the simulation engine can be used for generating graphical representations of the battlespace and the engaged operational units.

Using a PSF approach and taking advantage of HPC resources allows implementation of AI-assistive decision-making mechanisms exploiting big data ingests and analytics, while being available to geographically distributed users for collaborative decision-making efforts. A variety of mixed reality display modalities connected to a PSF-hosting server can support a range of operational scenarios from C2 at the strategic level to more mobile tactical use at the operational edge.

4.1.3 Realistic Representations of Decision Spaces

Graphically representing military decision-making strategies at all levels of operations requires new visualization approaches that can be applied to dynamic environments characterized by changing rules, cognitive states, uncertainty, and individual biases and heuristics.[17,44,45]

The visual representation of a battlespace should be as accurate and realistic as technologically possible, yet remain at a cognitive level that is humanly understandable and interpretable.[24,36,46,47] Advanced visualization approaches that incorporate mixed reality technologies have the potential to better represent the changing character of multi-domain warfare and its evolving threats and dynamic environments. With recent technological advancements in mixed reality visualization devices, lowered costs, and significantly improved reliability and usability of hardware, hybrid 2D and 3D visualization approaches are now possible.

Mixed reality approaches consisting of multiple 2D monitors augmenting more advanced 3D visualization capabilities can provide command staff with the necessary insights needed to understand complex war-gaming state spaces.[38] For example, the BVI[26] platform can realistically represent geospatial terrains using a combination of visualization modalities. As a data server, BVI distributes terrain, operational, and agent behavior data to client applications supporting multiple visualization modalities including Head-Mounted Display devices, web-based interfaces, mobile Android tablet devices, and mixed reality devices (e.g., HoloLens 2, Oculus Quest).

Figure 13 (top) shows a Friendly versus Hostile war-gaming scenario on BVI's high-resolution terrain of the Fort Irwin National Training Center located in San Bernardino County, California.[48] A 3D view of the battlespace can offer a more enriched user experience from multiple viewing perspectives than the traditional 2D map display often used during MDMP. The 3D view in BVI's Web Tactical Planner visualizes spatial information of both terrain and man-made features and the positions of the units depicted by MIL-STD-2525C symbols.[49]

Geospatial perspectives, such as those offered by BVI, conceivably support decision makers' understanding of dynamic battlespace environments. Paired with a navigable AI-augmented decision space (Fig. 13, bottom), the combined perspective can enable better understanding across visual-spatial dependencies, effects and causalities, estimated risks and values, uncertainty, and deception for complex decision-making. Combining such geospatial and decision-centric perspectives with AI may provide the necessary breadth to coordinate physical actions with actions in cyber and other nonspatial domains across multiple timescales, as well as the flexibility to adapt quickly to changing mission objectives and contexts.

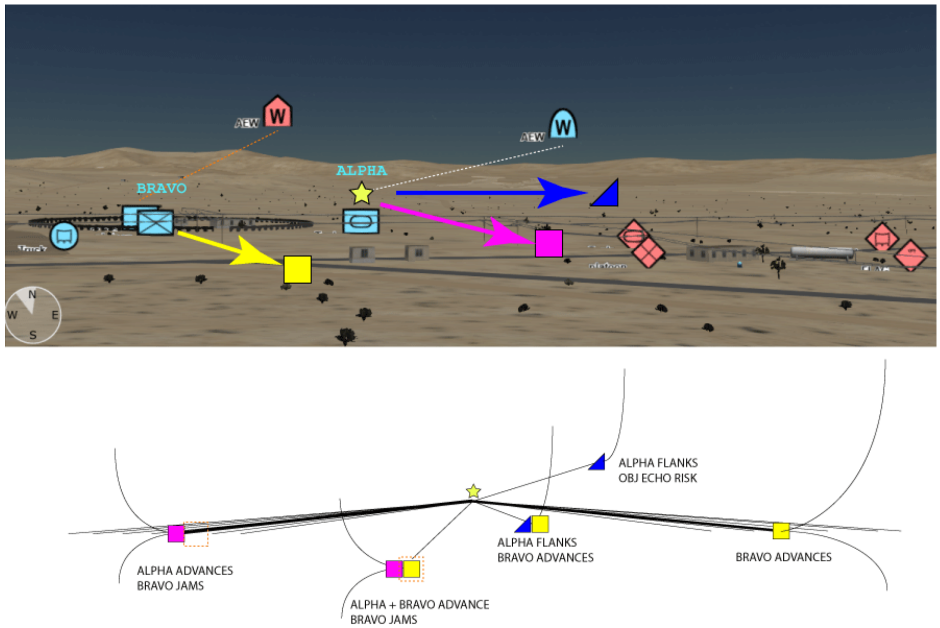

Fig. 13 3D view of a war-game scenario in the BVI Web Tactical Planner (top) with a concept of an AI decision tree (bottom)

5. Discussion

Opportunistic learning of naturalistic human decision-making behavior by AI,[42,50–52] and the appropriate structuring and sequencing of the learning environment so that the AI is judiciously shaped by the training curriculum,[53] are already established frameworks for improving the ability of AI to quickly learn difficult challenges. Further advancing the AI's capability for complex decision-making for war-gaming requires improving the ability of AI to tackle decisions under MDO contexts with high uncertainty, deception, and game theory, which are all challenges for reward assignment during AI development. Overcoming these challenges requires leveraging multidisciplinary advances ranging from neurobiological advances in understanding the brain's decision and reward circuitry and computations to the psychology of how expertise, tacit knowledge, theory of mind, game theory, and metareasoning are applied during complex decision-making.

How AIs can best learn from human complex decision-making remains an open question. Although the exact mechanisms for reward shaping for complex decision-making are yet to be discovered, this project has produced a vision of how to discover such mechanisms via a novel AI test-bed and WMIs. The ARL Battlespace AI test-bed and scenarios place the human and AI in the context of MDO-relevant decision-making, enabling the AI to learn how different decisions and factors interact and how humans navigate through such complex decision trees collaboratively and adversarially. A key advance is that the test-bed and scenarios provide a rich environment for efficiently developing AI Theory of Mind and MDO-relevant metareasoning for complex decision-making by abstracting away factors that would sparsify and impede learning of decision-making essentials.

Another advance is the development of high-performance computing frameworks to enable AI decision support for continuous distributed training. This will enable the hosting of the AI decision aids on ARL's Persistent Services Framework, so that in the future, Soldiers can train individually or collaboratively, in mixed human and AI teams, against an AI war-gaming agent anytime, anywhere.

A third advance of this project is the development of an approach for visualizing the AI's decision process to enable AI transparency and trust as well as collaborative human–AI team decision-making. The AI's reasoning must be both abstracted and relatable to the war-gaming environment, so that the human can understand the AI's valuation of different decision outcomes and efficiently navigate through the AI's decision tree without imposing undue cognitive burden. We have taken the first steps toward an AI-augmented WMI that is based on 3D mixed reality to harness and augment inherent human capacities for 3D cognition and prediction.[54–58] With further design, we envision that its interface will feel naturalistic while expanding the capability to display information from across multiple domains and enabling the AI to opportunistically learn from the user's decision-making. Such a naturalistic, intuitive AI-enabled decision aid, developed to support MDO C2 decision-making including deliberative and tacit reasoning, as well as collaborative and adversarial reasoning, is essential for humans to trust the AI estimations of COA outcomes in complex decision-making.

5.1 Potential for Further Development of the AI Test-Bed and AI Agents

While the exploitation of deep reinforcement learning algorithms in gaming has shown significant promise recently, a prerequisite of that success is working with a relatively simple, well-structured game. The real challenge arises as the environment increasingly relies on sparse observational data, complex and dynamic agent strategies, and sparse observational data. There are several tradeoffs to developing the platform entirely in house versus building on an existing open source library—primarily the minimization of constraints and sheer workload of environment development. Creating an entirely new custom platform allows for complete customization of game-related intricacies, albeit becoming very time consuming. Conversely, various impenetrable constraints emerge when using an existing library such as the StarCraft2LearningEnvironment (SC2LE),[59] but the work put into the game development decreases tenfold. Our ongoing second-generation development of the ARL Battlespace AI test-bed, named Simple Yeho (Fig. 14), is built upon a happy balance between both ends of the scale, OpenAI Gym,[60] a toolkit for developing reinforcement learning algorithms that makes no assumptions about the input agent and environment structure. A basic framework must be followed obviously, but OpenAI Gym provides the client complete design freedom in addition to a plethora of documentation and examples to follow. There are no immediate issues that need to be addressed from a game development point of view, but it does need to be a higher priority moving forward.

Fig. 14 Simple Yeho AI test-bed

Future issues are not limited to the game environment, as they will inevitably extend into theoretical reinforcement learning challenges such as seamless multiagent communication, task coordination, and stationary policies. More practical problems to focus on include algorithmic efficiency (limiting computationally intensive tasks as well as memory allocation mindfulness), a novel decentralized reinforcement learning algorithm, and data generalization across multiple domains. Overtaxing hardware resources is a common bottleneck in all branches of AI. From a software standpoint, the ARL Battlespace AI test-bed is quite frugal with its resources, and the environment remains focused on research questions for AI development rather than full MDO implementation, which is why computational efficiency is not yet a pressing issue. Potential solutions to generalizing game state information, particularly in a dynamic environment, include the temporal difference variational autoencoder[61] and distributional temporal difference reinforcement learning,[62] as they allow for explicit beliefs about several states in the future (which plays into the metareasoning aspect) in addition to providing a smooth latent space in between data points. Additional major problems our novel reinforcement learning algorithm should address are security/authentication, agent decision-making transparency, and real-time inter-agent communication. Integrating blockchain into the DEVCOM ARL framework would ensure secure lines of communication between nodes, provide an immutable and decentralized ledger to shed light on the agent's low-level decision making, and introduce a democratic voting system to the agents to promote group cooperation while still maintaining individual selfishness.

5.2 Potential for Further Development of the Human–AI Collaborative Interface

Limitations in the current military decision-making process defined a multidisciplinary research approach for development of a human and AI WMI for complex decision-making. A realistic representation of the decision space as a foundational layer consists of the battlespace terrain with geospatially accurate natural and man-made artifacts. An advanced yet intuitive user interface allowing mixed reality perspectives of the battlespace enables decision-maker exploration of COA alternatives based on operational factors. Both these requirements guided selection of an Army and commercially developed battlespace interaction system, BVI, as a potential transition medium for the AI and human-AI teaming development achieved in the ARL Battlespace AI test-bed.

An initial step toward transition involved overlaying the grid-like environment of ARL Battlespace on a BVI real-world operational terrain and adapting the existing BVI multimodal user interface for war-gaming. Figure 15 shows a section of an expanded grid overlayed on Fort Irwin terrain using BVI's Web Tactical Planner 3D viewer perspective with friendly and hostile units situated at the start of a war-gaming session. Units can be placed and manipulated using a tactical planning toolbar with mouse, trackpad, or touchscreen interaction within a browser window. BVI provides the capabilities to add units; route points, tactical symbols, and graphics; and draw features such as lines, polygons, and text boxes.

Fig. 15 3D view of a war-game scenario in the BVI Web Tactical Planner with grid overlay

An unresolved question is how best to harness BVI's mixed reality (XR) visualization capabilities for collaborative decision-making (e.g., by sharpening decision-makers' understanding of the terrain's geospatial factors during war-gaming). The ability to load different terrains and create customized training scenarios potentially derived from multidimensional data and viewed in a variety of immersive formats exceeds visualization capabilities of other Army systems. Depending on the expanse and detail of these 3D terrains, how the interface displays this information could cause substantial information overload or confusion as the decision-maker maneuvers throughout large areas of terrain using a robust array of interaction modalities. An effective interface would need to be designed to select not only what environmental and decision space information to communicate, but also how to present that information from the user's vantage point.

If development time and effort are not possible, BVI's API offers opportunities to embed visual aids in the form of markings, labels, and scenario-adaptive grids positioned on top of the terrain as spatial management interventions for decision-makers. For example, the rows and columns of the grid depicted in Fig. 15 could be labeled or coded for quickly locating real-time events and AI-generated activities. Multidimensional gridded structures and coding schemes could elevate war-gaming to a level of complexity characterized by MDO while mitigating some of the terrain-based spatial management concerns.

Coordinating multiple views of the battlespace with data anlysis in both spatial and temporal domains, visualizations provide for additional approaches that facilitate complex decision-making during war-gaming. When a shared MDO battlespace representation is required, a collaborative strategic planning mode can be achieved with multiple coordinated views implemented on different visualization modalities to update interactively based on distributed command staff inputs. Command staff inputs can also guide the application of visual filters to the coordinated views, resulting in a reduction of unnecessary complexity and an accentuation of scenario or mission-critical battlespace information.

Figure 16 shows the SyncVis[63] visual analytics system designed to display multiple coordinated views of data analysis supporting data exploration and understanding. SyncVis generates multiple data representations by linking the information shown in each view to the others via user interactions. This example shows SyncVis's 2D interface for a COVID categorical population data analysis in four coordinated views: Variable Selector (six attributes selected), Map/Terrain, Mutual Information Diagram, and Stacked Area Chart for each selected variable.

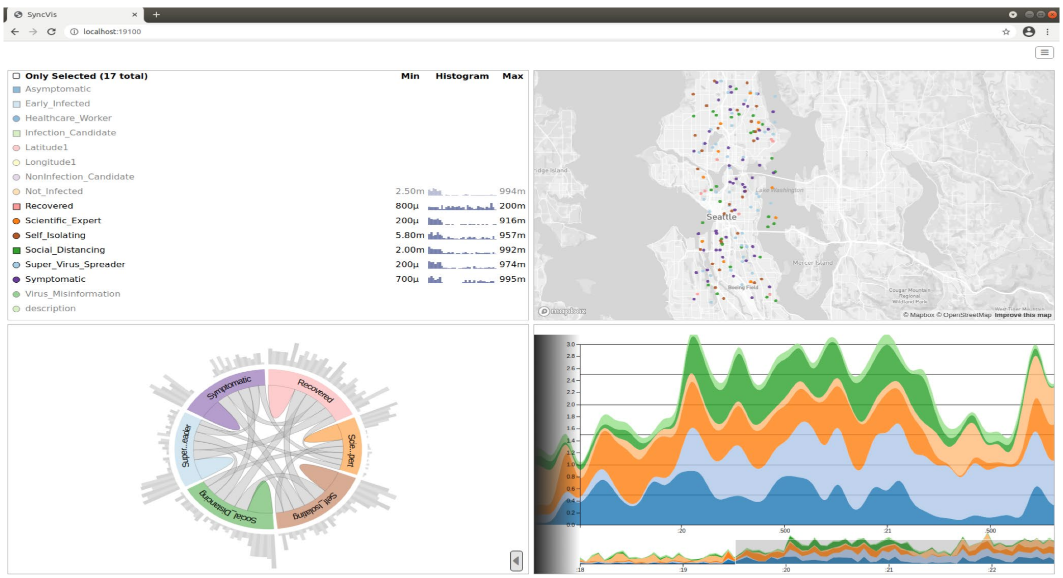

Fig. 16 **SyncVis 2D interface showing multiple coordinated visualizations of COVID data analyses**

The SyncVis visualization capabilities can be integrated with an HPC analytic workflow back end using PSF. The PSF server can stream operational and agent behavior data to both BVI and SyncVis, creating a unified battlespace exploration experience.[64] The benefits of coordinating battlespace views based on user on-demand input and filtering are open for investigation.

A flexible war-gaming environment appears to be key as each training scenario, COA, and mission plan is developed within the constraints of the MDMP and associated military doctrine but is uniquely different and dependent on the battlespace and its operational variables. An HPC PSF data analysis processing pipeline powering a WMI with Soldier or commander on-demand coordinated BVI and SyncVis visualizations of the battlespace would revolutionize the existing war-gaming paradigm and touch the level of complexity inherent in MDO and the level of human and AI-directed decision guidance required to win.

6. Conclusion

We have highlighted three key areas for development: the AI-directed decision guidance,[24,29,36,65] the computational infrastructure to support this guidance, and the development of mixed reality representations for decision transparency.[22,46] Advances in these areas require expertise across many different disciplines. Novel AI development requires the fusion of ideas from neuroscience, psychology, and mathematics to overcome bottlenecks to long-standing problems in complex decision-making. This includes learning across long time scales and catastrophic forgetting under changing contexts, as well as problems more specific to war-gaming, such as multiagent decision-making with uncertainty, deception, and game theory. The computational infrastructure also needs development, as computing power and data frameworks are both essential for producing common operating pictures for human–AI teaming at the tactical edge. For efficient development, proprietary restrictions and software dependencies should be abstracted away via a common framework, with clear documentation for usage and troubleshooting to allow academia, government, and industry to better focus on tackling the human–AI teaming problem. This common framework should include efficient passing of information while providing flexibility and adaptability to the needs of both the AI development and the human user across both training and live-use environments. Finally, the development of the interface itself needs concerted expertise across multiple disciplines. A foundational problem is how to compact information to be efficiently understood by the user, and how to best harness user interactions for opportunistic learning. The human mind does not process all sensory information, but instead makes predictions and assumptions about the world to economize its computations under an environment with incomplete information. An effective WMI should anticipate both potential decision outcomes as well as individual user expectations and assumptions, and it should provide alerting and suggest counterstrategies for possible friendly and hostile COAs and deception. Additionally, the AI decision aid must estimate what is the user's tacit understanding, allowing it to present the most relevant information and the most promising choices pulled from across the Warfighting domains.

7. References

1. Allen G. Understanding AI technology. Joint Artificial Intelligence Center (JAIC), the Pentagon; 2020.
2. Vorm ES. Computer-centered humans: why human-AI interaction research will be critical to successful AI integration in the DoD. IEEE Intelligent Systems. 2020;35(4):112–116.
3. Feickert A. Defense primer: Army multi-domain operations (MDO). Congressional Research SVC; 2021.
4. Center for Army Lessons Learned. Military decisionmaking process: lessons and best practices. Center for Army Lessons Learned. US Army Combined Arms Center; 2015.
5. Bailey K. Commanders plan; autonomy facilitates; DEVCOM C5ISR Center News and Media; 2017 [accessed 2021 Dec 9]. https://www.army.mil/article/193398/commanders_plan_autonomy_facilitates.
6. Surdu JR, Haines GD, Pooch UW. OpSim: a purpose-built distributed simulation for the mission operational environment. Simulation Series. 1999;31:69–74.
7. West CTD, Birkmire B. AFSIM: The Air Force Research Laboratory's approach to making M&S ubiquitous in the weapon system concept development process. Cybersecurity & Information Systems Information Analysis Center. 2020;7(3):50.
8. PEO_STRI. One Semi-Automated Forces (OneSAF); 2022 [accessed 2022 Feb 20]. https://www.peostri.army.mil/OneSAF
9. Silver D, Hubert T, Schrittwieser J, Antonoglou I, Lai M, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. Science. 2018;362(6419):1140–1144.
10. Gamage C, Pinto V, Xue C, Stephenson M, Zhang P, Renz J. Novelty generation framework for AI agents in Angry Birds style physics games. 2021 IEEE Conference on Games (CoG). 2021;1–8.
11. Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, Riedmiller M. Playing Atari with deep reinforcement learning; 2013. arXiv preprint arXiv:1312.5602.

12. Goecks VG, Waytowich N, Asher DE, Park SJ, Mittrick M, et al. On games and simulators as a platform for development of artificial intelligence for command and control; 2021. arXiv preprint arXiv:2110.11305.

13. Goecks VG, Waytowich N, Watkins D, Prakash B. Combining learning from human feedback and knowledge engineering to solve hierarchical tasks in Minecraft; 2021. arXiv preprint arXiv:2112.03482.

14. Narayanan P, Wu Z, Kwon H, Wang Z, Rao R. Overview of machine learning (ML) based perception algorithms for unstructured and degraded visual environments. In: SPIE Defense + Commercial Sensing: Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications. International Society for Optics and Photonics; 2019.

15. Köster R, Hadfield-Menell D, Everett R, Weidinger L, Hadfield GK, Leibo JZ. Spurious normativity enhances learning of compliance and enforcement behavior in artificial agents. Proceedings of the National Academy of Sciences. 2022;119(3).

16. Hare JZ, Uribe CA, Kaplan LM, Jadbabaie A. Communication constrained learning with uncertain models. ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2020;8609–8613.

17. Raglin A, Dennison M, Metu S, Trout T, James D. Decision making with uncertainty in immersive systems. In: Virtual, augmented, and mixed reality (XR) technology for multi-domain operations. International Society for Optics and Photonics; 2020. 114260L.

18. Lavine M. WHIM: function approximation where it matters. Communications in Statistics-Simulation and Computation. 2019;1–31.

19. Chen JY, Flemisch FO, Lyons JB, Neerincx MA. Guest editorial: agent and system transparency. IEEE Transactions on Human-Machine Systems. 2020;3(50):189–193.

20. Chen JY, Lakhmani SG, Stowers K, Selkowitz AR, Wright JL, Barnes M. Situation awareness-based agent transparency and human-autonomy teaming effectiveness. Theoretical Issues in Ergonomics Science. 2018;19(3):259–282.

21. Rinderspacher BC. Smooth constrained heuristic optimization of a combinatorial chemical space. Army Research Laboratory (US); 2015. Report No.: ARL-TR-7294.

22. Su S, Kase S, Hung C, Hare JZ, Rinderspacher BC, Amburn C. Mixed reality visualization of friendly vs hostile decision dynamics. In: International Conference on Human-Computer Interaction. Mixed Reality Technology for Immersive Analytics. Springer; 2021.

23. Mait JN. A report on Army science planning and strategy 2016. Army Research Laboratory (US); 2017. Report No.: ARL-SR-0375.

24. Hare JZ, Rinderspacher BC, Kase SE, Su S, Hung CP. Battlespace: using AI to understand friendly vs. hostile decision dynamics in MDO. SPIE Defense + Commercial Sensing: Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications III; 2021. p. 1174615; Vol. 11746. https://doi.org/10.1117/12.2585785.

25. Cirincione G, Verma D. Federated machine learning for multi-domain operations at the tactical edge. Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications; 2019. p. 1100606; Vol. 11006.

26. Garneau CJ, Boyce MW, Shorter PL, Vey NL, Amburn CR. The Augmented Reality Sandtable (ARES) research strategy. Army Research Laboratory (US); 2018. Report No.: ARL-TN-0875.

27. Joint Staff. Charter of the Joint Requirements Oversight Council (JROC) and Implementation of the Joint Capabilities Integration and Development System (JCIDS); 2018. CJCSI 5123.01H.

28. Headquarters, Department of the Army. Offense and defense; 2019. Army Doctrine Publication (ADP) 3-90.

29. Hare JZ, Rinderspacher BC, Peregrim W, Kase SE, Su SM, Hung CP. ARL Battlespace; 2021 [accessed 2022 Mar 20]. https://github.com/USArmyResearchLab/ARL_Battlespace.

30. Rabinowitz N, Perbet F, Song F, Zhang C, Eslami SA, Botvinick M. Machine theory of mind. International Conference on Machine Learning. 2018;4218–4227.

31. Karimi S, Hamidi G, Fatahi Z, Haghparast A. Orexin 1 receptors in the anterior cingulate and orbitofrontal cortex regulate cost and benefit decision-making. Progress in Neuro-Psychopharmacology and Biological Psychiatry. 2019;89:227–235.

32. Delgado-Herrera M, Reyes-Aguilar A, Giordano M. What deception tasks used in the lab really do: systematic review and meta-analysis of ecological validity of fMRI deception tasks. Neuroscience. 2021;468:88–109.

33. Abe N, Greene JD, Kiehl KA. Reduced engagement of the anterior cingulate cortex in the dishonest decision-making of incarcerated psychopaths. Social Cognitive and Affective Neuroscience. 2018;13(8):797–807.

34. Gardner JL. Optimality and heuristics in perceptual neuroscience. Nature Neuroscience. 2019;22(4):514–523.

35. Tiippana K. What is the McGurk effect? Frontiers in Psychology. 2014;5:725.

36. Hung CP, Hare JZ, Rinderspacher BC, Peregrim W, Kase S, Su S, Raglin A, Richardson JT. ARL Battlespace: a platform for developing novel AI for complex adversarial reasoning in MDO. In: SPIE Defense + Commercial Sensing (SI22). Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications IV (Conference 12113); 2022.

37. Gilmore JM. Nett warrior initial operational test and evaluation (IOT&E) report. Director, Operational Test and Evaluation; 2015.

38. Su S, Kase S, Hung C, Hare JZ, Rinderspacher BC, and Amburn C. Mixed reality visualization of friendly vs hostile decision dynamics. International Conference on Human-Computer Interaction. 2021;545–555.

39. Drubin C. AlphaDogfight trials foreshadow future of human-machine symbiosis. Horizon House Publications; 2020.

40. Grüning DJ, Krueger JI. Strategic thinking: a random walk into the rabbit hole. Collabra: Psychology. 2021;7(1):24921.

41. Cassenti DN, Kaplan LM. Robust uncertainty representation in human-AI collaboration. Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications III; 2021. p. 1174613; Vol. 11746.

42. Lance BJ, Larkin GB, Touryan JO, Rexwinkle JT, Gutstein SM, et al. Minimizing data requirements for Soldier-interactive AI/ML applications through opportunistic sensing. SPIE Defense + Commercial Sensing: Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications II; 2020. p. 1141306; Vol. 11413.

43. Gil Y, Pierce SA, Babaie H, Banerjee A, Borne K, et al. Intelligent systems for geosciences: an essential research agenda. Communications of the ACM. 2018;62(1):76–84.

44. Dennison M, Reardon C, Gregory J, Trout T, Rogers III JG. Creating a mixed reality common operating picture across C2 echelons for human-autonomy teams. Virtual, Augmented, and Mixed Reality (XR) Technology for Multi-Domain Operations; 2020. p. 114260M; Vol. 11426.

45. Hung CP, Callahan-Flintoft C, Fedele PD, Fluitt KF, Odoemene O, et al. Abrupt darkening under high dynamic range (HDR) luminance invokes facilitation for high contrast targets and grouping by luminance similarity. J Vision. 2020;20:9.

46. Kase SE, Hung CP, Krayzman T, Hare JZ, Rinderspacher BC, Su SM. The future of collaborative human-AI decision-making for mission planning. Frontiers in Psychology; 2022 Apr 4. DOI: 10.3389/fpsyg.2022.850628.

47. Larkin GB, Geuss MN, Yu A, Callahan-Flintoft C, Rexwinkle J, Hung CP, Lance BJ. Human performance with complex technology: how visual cognition is critical to enhanced performance with aided target recognition (AiTR). International Conference on Applied Human Factors and Ergonomics; 2020. p. 134–140.

48. Fort Irwin Training Center. Wikipedia; last updated 2022 Mar 21 [accessed 2021 Nov 9]. https://en.wikipedia.org/wiki/Fort_Irwin_National_Training_Center.

49. MIL-STD-2525D. Department of Defense interface standard. Joint military symbology. Department of Defense; 2014 Jun 10.

50. Hung CP, Choi J, Gutstein SM, Jaswa MS, Rexwinkle JT. Soldier-led adaptation of autonomous agents (SLA3). Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications III; 2021. p. 117462U; Vol. 11746.

51. Cox KR, Rexwinkle JT, Gremillion GM, Perelman BS, Pollard KA, et al. A summary of the effects of technology aids and crew size on a next generation combat vehicle platoon section. DEVCOM Army Research Laboratory; 2022. Report No.: ARL-TN-1109.

52. Cox KR, Rexwinkle JT, Gremillion GM, Perelman BS, Brooks JR, et al. Effects of technology aids and crew size on a Next Generation Combat Vehicle platoon section. DEVCOM Army Research Laboratory; 2022. Report No.: ARL-TR-9403.

53. Warnell G, Waytowich N, Lawhern V, Stone P. Deep tamer: interactive agent shaping in high-dimensional state spaces. Proceedings of the AAAI Conference on Artificial Intelligence; 2018.

54. Boyce MW, Gardony AL, Shorter P, Horner C, Stainrod CR, Flynn J, Brunyé TT, Amburn CR. Characterizing the cognitive impact of tangible augmented reality. International Conference on Human-Computer Interaction; 2019. p. 416–427.

55. Kamitani Y, Tong F. Decoding seen and attended motion directions from activity in the human visual cortex. Current Biology. 2006;16(11):1096–1102.

56. Kim H, Lee S, Bovik AC. Saliency prediction on stereoscopic videos. IEEE Transactions on Image Processing. 2014;23(4):1476–1490.

57. Krokos E, Plaisant C, Varshney A. Virtual memory palaces: immersion aids recall. Virtual Reality. 2019;23(1):1–15.

58. Welchman AE, Deubelius A, Conrad V, Bülthoff HH, Kourtzi Z. 3D shape perception from combined depth cues in human visual cortex. Nature Neuroscience. 2005;8(6):820–827.

59. Vinyals O, Ewalds T, Bartunov S, Georgiev P, Vezhnevets AS, et al. Starcraft II: a new challenge for reinforcement learning; 2017. arXiv preprint arXiv:1708.04782.

60. Brockman G, Cheung V, Pettersson L, Schneider J, Schulman J, Tang J, Zaremba W. OpenAI gym; 2016. arXiv preprint arXiv:1606.01540.

61. Gregor K, Papamakarios G, Besse F, Buesing L, Weber T. Temporal difference variational auto-encoder; 2018. arXiv preprint arXiv:1806.03107.

62. Dabney W, Kurth-Nelson Z, Uchida N, Starkweather CK, Hassabis D, Munos R, Botvinick M. A distributional code for value in dopamine-based reinforcement learning. Nature. 2020;577(7792):671–675.

63. Su S, Perry V, Bravo L, Kase S, Roy H, Cox K, Dasari VR. Virtual and augmented reality applications to support data analysis and assessment of science and engineering. Computing in Science & Engineering. 2020;22(3):27–39.

64. Syncvis; 2022 [accessed 2022 Mar 22. https://gitlab.hpc.mil/simon.m.su.civ/syncvis.

65. Caylor J, Herrmann JW, Hung C, Raglin A, Richardson J. Metareasoning for multi-criteria decision making using complex information sources. In. SPIE Defense + Commercial Sensing; 2022.

Appendix. Learning from Historical Data

Let the set of actions of a unit i be defined as $A^i = \{a_1^i, \dots, a_{K^i}^i\}$ and the set of possible states of the unit be defined as $S^i = \{s_1^i, \dots, s_M^i\}$, where M is the total number of states of all of the units. Following from the previous section, the probability of an action is modeled as a categorical distribution

$$P\left(a_k^i \middle| s_m^i\right) = \prod_{x=1}^{K^i} \pi^i\left(x \middle| s_m^i\right)^{[x=k]}$$

where $\pi^i\left(x \middle| s_m^i\right)$ is the probability of taking an action x given that the unit i is in state s_m^i. One approach to modeling these probabilities is to use the frequentist approach as taken in the previous section. However, due to limited data, the epistemic uncertainty in the probabilities $\pi^i\left(x \middle| s_m^i\right)$ must be incorporated, as some of the action state pairs may not be realized in the overall set of game data.

Following a Bayesian approach, we can account for the epistemic uncertainty in the probabilities by utilizing a prior on the distribution of π^i. To identify the prior, we first analyze the game data D and compute the frequency of times a unit has taken an action while in a particular state, i.e., identify the set $N^i = \left\{N\left(a_1^i \middle| s_1^i\right), \dots, N\left(a_1^i \middle| s_M^i\right), \dots, N\left(a_{K^i}^i \middle| s_M^i\right)\right\}$, which is a set of counts that represent the number of times an action was taken while the unit was in a particular state. Then, the distribution of the probabilities π^i are modeled according to a Dirichlet distribution

$$f\left(\pi^i \middle| s_m^i, D, \alpha\right) = \frac{1}{B(\alpha)} \prod_{x=1}^{K^i} \pi^i\left(x \middle| s_m^i\right)^{\alpha_x - 1 + N(a_x^i|s_m^i)},$$

where $\alpha = \{\alpha_1, \dots, \alpha_{K^i}\}$ are the set of hyperparameters s.t. $\alpha_x > 0$ and $B(\alpha) = (\prod_{x=1}^{K^i} \Gamma(\alpha_x))/\Gamma(\sum_{x=1}^{K^i} \alpha_x)$ is the multivariate beta function. We model the distribution of probabilities as a Dirichlet distribution since it is the conjugate prior of a categorical distribution. Following the Bayesian bootstrap[1] and imprecise Dirichlet models,[2] the hyperparameters were chosen to be sufficiently close to 0 to allow the human data to dictate the actions taken by the random agent. Future work will analyze the optimal choice of the hyperparameters given the defined model.

Now, since the underlying probabilities of each action conditioned on the state are unknown, we compute the probability of an action as the posterior predictive distribution:

---

[1] Rubin DB. The Bayesian bootstrap. The Annals of Statistics. 1981;130–134.
[2] Bernard J-M. An introduction to the imprecise Dirichlet model for multinomial data. International Journal of Approximate Reasoning. 2005;39(2–3):123–150.

$$P\left(a_k^i \middle| s_m^i, D\right) = \int_{\pi^i} P\left(a_k^i \middle| s_m^i, \pi^i\right) f\left(\pi^i \middle| s_m^i, D, \alpha\right) d\pi^i = \frac{N\left(a_k^i \middle| s_m^i\right) + \alpha_k + 1}{\sum_{x=1}^{K^i}\left(\alpha_x + N\left(a_x^i \middle| s_m^i\right)\right) + K^i} \tag{A-1}$$

This distribution is the expected value of the action probabilities conditioned on the posterior Dirichlet distribution conditioned on the game data. A key property that comes from using the posterior predictive distribution instead of the frequency distribution is that the probability of an action state pair is always greater than 0, even if the action state pair was not realized in the game data.

In general, we can also compute the probabilities of joint actions and states following the same approach as

$$P\left(a_k^i, a_{k'}^j \middle| s_m^i, s_{m'}^i, D\right) = \frac{N\left(a_k^i, a_{k'}^j \middle| s_m^i, s_{m'}^i\right) + \alpha_{kk'} + 1}{\sum_{x=1}^{K^i K^j}\left(\alpha_x + N\left(a_x^i, a_{x'}^j \middle| s_m^i, s_{m'}^i\right)\right) + K^i K^j}$$

where $N\left(a_k^i, a_{k'}^j \middle| s_m^i, s_{m'}^i\right)$ are the number of times the joint actions and states were taken and $\alpha_{kk'}$ is the hyperparameter for the joint actions and states pair. Implementing the joint actions into the overall framework will be studied as future work.

List of Symbols, Abbreviations, and Acronyms

2D	two-dimensional
3D	three-dimensional
AI	Artificial Intelligence
AFSIM	Advanced Framework for Simulation, Integration and Modeling
APF	Automated Planning Framework
API	application programming interface
ARES	Augmented REality Sandtable
ARL	Army Research Laboratory
BVI	Battlespace Visualization and Interface (formerly ARES)
C2	command and control
COA	course of action
DEVCOM	US Army Combat Capabilities Development Command
DFV	Director's Future Ventures
DOD	Department of Defense
DOTMLPF	doctrine, organization, training, materiel, leadership and education, personnel, facilities
HPC	high-performance computing
MDMP	military decision-making process
MDO	Multi-Domain Operations
METT-TC	mission, enemy, terrain & weather, troops, time available, and civil considerations
MITM	man-in-the-middle
OneSAF	One Semi-Automated Forces
OpSim	Operationally focused simulation tool
PSF	Persistent Services Framework
S&T	science and technology

SC2LE	StarCraft 2 Learning Environment
Simple Yeho	“Simple Yehorivka” AI test-bed
WMI	Warfighter machine interface

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

1 DEVCOM ARL
(PDF) FCDD RLD DCI
TECH LIB

1 DEVCOM ARL
(PDF) FCDD RLH B
T DAVIS
BLDG 5400 RM C242
REDSTONE ARSENAL AL
35898-7290

1 DEVCOM ARL
(PDF) FCDD HSI
J THOMAS
6662 GUNNER CIRCLE
ABERDEEN PROVING
GROUND MD
21005-5201

1 USN ONR
(PDF) ONR CODE 341 J TANGNEY
875 N RANDOLPH STREET
BLDG 87
ARLINGTON VA 22203-1986

1 USA NSRDEC
(PDF) RDNS D D TAMILIO
10 GENERAL GREENE AVE
NATICK MA 01760-2642

1 OSD OUSD ATL
(PDF) HPT&B B PETRO
4800 MARK CENTER DRIVE
SUITE 17E08
ALEXANDRIA VA 22350

1 DA HQ
(PDF) DASA(R&T)

9 USARMY AFC
(PDF) L BROUSSEAU
J REGO
A LINZ
K WADE
S BRADY
J REGO
T KELLY
E JOSEPH
B SESSLER

2 DEVCOM HQ
(PDF) FCDD ST
C SAMMS
M HUBBARD

2 OAK RIDGE ASSOCIATED
(PDF) UNIVERSITIES
W PEREGRIM
T KRAYZMAN

1 NIST
(PDF) S SU

ABERDEEN PROVING GROUND

89 DEVCOM ARL
(PDF) FCDD RLC
C BEDELL
B SADLER
B PIEKARSKI
H EVERITT
FCDD RLC CA
J HARE
L KAPLAN
FCDD RLC ES
G VIDEEN
S HILL
Y PAN
FCDD RLC I
B MACCALL
FCDD RLC N
BM RIVERA
A SWAMI
FCDD RLC IT
A RAGLIN
J RICHARDSON
FCDD RLC NC
S KASE
FCDD RLC S
O TKACHENKO
FCDD RLD
P BAKER
A KOTT
S SILTON
FCDD RLD D
T ROSENBERGER
FCDD RLD E
KS FOSTER
FCDD RLD F
K KAPPRA
FCDD RLD FR
M TSCHOPP
FCDD RLD SM
L BLUM
FCDD RLH

J LANE
Y-S CHEN
P FRANASZCZUK
K MCDOWELL
FCDD RLH B
JJ SUMNER
FCDD RLH F
J GASTON
K OIE
FCDD RLH FA
AW EVANS
G BOYKIN
FCDD RLH FB
J GARCIA (A)
H ROY
FCDD RLH FC
J TOURYAN (A)
T ROHALY
C HUNG
W PEREGRIM
T KRAYZMAN
FCDD RLH FD
A MARATHE
FCDD RLH T
D STRATIS-CULLUM
FCDD RLL
T KINES
FCDD RLL D
J S ADAMS
FCDD RLL DP
J MCCLURE
FCDD RLR
B HALPERN
S LEE
D STEPP
FCDD RLR E
RA MANTZ
C VARANASI
FCDD RLR EL
JX QIU
MD ULRICH
FCDD RLR EN
RA ANTHENIEN JR
FCDD RLR IC
MA FIELDS
SP IYER
FCDD RLR IM
JD MYERS
FCDD RLR IN
XN WANG
FCDD RLR
P REYNOLDS
FCDD RLR P
LL TROYER
FCDD RLR PC
D POREE
FCDD RLR PL
MK STRAND
FCDD RLS
J ALEXANDER
M GOVONI
M WRABACK
FCDD RLS C
M REED
FCDD RLS CC
S BEDAIR
FCDD RLS CE
TR JOW
K XU
FCDD RLS E
RD DELROSARIO
FCDD RLS ED
K JONES
FCDD RLS EA
A ZAGHLOUL
FCDD RLS S
WL BENARD
FCDD RLS SO
W ZHOU
FCDD RLW
S KARNA
JF NEWILL
AM RAWLETT
SE SCHOENFELD
J ZABINSKI
FCDD RLW B
R BECKER
FCDD RLW M
ES CHIN
FCDD RLW S
V CHAMPAGNE
AL WEST
FCDD RLW T
RZ FRANCART
FCDD RLW TC
JD CLAYTON
FCDD RLW W
TV SHEPPARD
FCDD RLW WA
B RICE
R PESCE-RODRIGUEZ
FCDD RLW M
A HALL
FCDD RLW MC
B RINDERSPACHER