Technical Evaluation Report

Katherine L. Morse, PhD
Principal Professional Staff
The Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Road, Laurel, MD, USA 20723

katherine.morse@jhuapl.edu

Curtis L. Blais, PhD
Research Faculty
MOVES Institute, Naval Postgraduate School
1 University Circle, Monterey, CA, USA, 93943

clblais@nps.edu

EXECUTIVE SUMMARY

The NATO Modelling and Simulation Group (NMSG) Conference (MSG-207), "Simulation: Going Beyond the Limitations of the Real World," was conducted in Monterey, CA, USA on 19 & 20 October 2023. All sessions of the Conference were unclassified. The Conference audience included participants from NATO countries, Partners-for-Peace (NP) nations, and invited nations. At the Conference, 19 papers were presented. A keynote presentation, an invited presentation, and a briefing from the President of the Naval Postgraduate School were given on the first day. An additional invited presentation was given on the second day. Additionally, a panel session and two poster sessions were presented on the first day. Conference contributions were presented throughout in nine topic sessions. This Technical Evaluation Report summarizes the core ideas and results presented in this conference. This report also provides an overview of discussions conducted during the conference and concludes with a summary and recommendations section.

The theme for MSG-207 was the next generation synthetic battlespace addressing current limitations in training (e.g., resource availability, information protection); simulation for decision support (e.g., providing broader and deeper analysis of complex mission area issues); and Concept Development and Experimentation (CD&E; e.g., providing a broader and deeper analysis of new concepts and systems that are not yet available such as hypersonic capabilities, directed energy weapons, and swarming Unmanned Vehicles (UxVs) (air, ground, surface, subsurface) etc.).

Papers were invited to discuss advances in M&S technologies needed to achieve these goals, including cloud computing, AI and Machine Learning, Extended Reality (XR) or the Military Metaverse, metrics and standards across the full range of application domains: including air, land, maritime, cyber, space, autonomous systems, influence and information operations, C4ISR (Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance), and military-civil decision-making support. Papers were also recommended to discuss technology and processes needed to create eco-systems where M&S service providers and consumers can meet to discover, compose, and deliver the capabilities that are needed by the warfighter.

Topics included CD&E in Session 1, training effectiveness in Session 2, decision support in NATO Federated Mission Networking in Session 3, architectures and interoperability supporting distributed training and test and evaluation (T&E) in Session 4, decision support more generally in Session 5, adapting M&S for national

needs in Session 6, digital twins in Session 7, artificial intelligence (AI) / machine learning (ML) in Session 8, and human behaviour representation (HBR) in Session 9.

Summary recommendations include:

1. *External to the symposium itself, value may be provided to participants through a virtual mechanism, such as a wiki or knowledge base, for sharing results of ongoing work.*

2. *Collect and consolidate NATO uses of M&S standards.*

3. *Consider establishment of an exploratory team or specialist team to develop a white paper to explore the intersection of M&S domains, use cases for each domain that could benefit from AI/ML, specific AI/ML most applicable to the use cases, and risks and challenges associated with applying specific types AI/ML to the use cases*

4. *Consider as a future theme rearchitecting systems to integrate emerging technologies, such as AI / ML, Cyber / Electronic Warfare (EW) and HBR.*

1.0 INTRODUCTION

The theme for MSG-207 was the next generation synthetic battlespace addressing current limitations in training, simulation for decision support, and Concept Development and Experimentation (CD&E), e.g., providing a broader and deeper analysis of new concepts and systems that are not yet available such as hypersonic capabilities, directed energy weapons, and swarming Unmanned Vehicles (UxVs) (air, ground, sea).

This Technical Evaluation Report contains a synopsis of the plenary and paper presentations. Also included are significant technical and thematic aspects of the conference, concluding with recommendations for consideration by the organizers. While this Technical Evaluation Report serves as a record of the conference's activities, the best source of the information presented are the original papers and the related presentation materials posted to the Science and Technology Organization (STO) ScienceConnect MSG-207 site.

2.0 DAY 1 (THURSDAY, 19 OCTOBER 2023)

2.1 Invited Presentation: Going Beyond the Limitations of the Real World; Mr Ryan MCALINDEN, The Institute for Creative Technologies

Mr MCALINDEN's presentation focused on orientation of M&S techniques, technologies, and research areas. How can we apply these to new problems across the armed forces? The presenter provided an overview of USC's Institute for Creative Technologies (ICT), describing immersive experiences that leverage research technologies, the art of entertainment, and storytelling to simulate the human experience to benefit learning, education, health, operations, human performance, and knowledge. Technologies that apply to creating these immersive experiences include:

- Artificial intelligence
- Virtual reality
- 3D terrain
- Computer graphics

Mr MCALINDEN reviewed the history of ICT going back to 1997 and identified the organization's core competencies as:

- Artificial intelligence
- Learning sciences
- Integration technology
- Human performance, optimization, and health
- Creative technologies: graphics, immersion, virtual humans, and narrative

Recent writers' and actors' strikes highlight the impact of AI on Los Angeles' creative work force. The presenter highlighted the challenges of explainability of AI/ML and described the relationship between the Star Trek holodeck and how technology is improving in the direction of that concept.

Major science and technology (S&T) areas that intersect for operations, training, simulation, mission command, and 3D geospatial terrain include:

- Environmental – how we can create representations of our environment to operate across communities, e.g., new mission command platforms being built in the way commercial video games are built
- Vision and graphics
- Natural language dialogue
- Virtual humans / AI – seeing generation of full, virtual humans rather than just animating avatars from the head up; looking for solutions not requiring a graphics artist
- Affective computing – how elements of cognition work along with humans, e.g., negotiation; they're working to shift from research to applications
- Multimodal perception – gesture and motion determination feeding into analytics, which can apply to intel and training, i.e., by seeing and understanding how individuals perform tasks, interact with systems, and react to training
- Body computing for health and human performance – understanding and tracking team performance

Mr MCALINDEN also discussed operationalization through integration, which is the intersection of multi-modal perception, virtual human / language dialogue, learning science, and mixed reality as "creative arts + subject matter experts (SMEs) + engineering" to create training applications. The automation space is taking over manual tasks, e.g., purely synthetic generation of 3D terrain. But are the results real enough to achieve the desired end goal? The presenter described the strategic advantage of imagination having the goal of leveraging scientists, engineers, military leaders, and Hollywood creative talent to discuss, ideate, and envision future warfare possibilities. So, how do you rate the power of storytelling and future storytelling? Writers see problem spaces differently than scientists, engineers, and military SMEs do. Long range fires have benefited from interactive sessions with writers and graphic artists. The presenter expressed concern that all domain operations have "COIN hangover," i.e., the younger cohort has little knowledge and no experience of large-scale, division/Corps-level tactics, techniques, and procedures (TTPs), and 80% have no near-peer engagement experience. There is a danger of losing hard-earned lessons from the years of counter-insurgency operations.

2.2 Keynote Speech: Mr Richard TEMPALSKI, Department of the Air Force Chief Modelling & Simulation Officer (CMSO)

Mr TEMPALSKI's presentation explained CMSO's establishment and position within the US Air Force (USAF). All aspects of M&S are under CMSO's purview. The US Air Force (USAF) spends \$3-\$4B / year on M&S. They wrote a policy requiring each M&S organization to provide one person to contribute to pain points and set future M&S direction for 14 areas, which include:

- Standards
    - Simulation interoperability
    - Use of SysML

- Medical
- Wargaming
- Augmented virtual worlds
- Logistics
- Infrastructure, e.g., for repositories and license servers

Mr TEMPALSKI's office is digitizing text manuals and using augmented reality (AR) to train crew chiefs in 3-5 years instead of 7-10 years and saving millions of dollars. The presenter is also the chief of AI responsible for validating AI algorithms developed by others and assessing the ethics of employment of those algorithms. Workforce training falls within Mr TEMPALSKI's purview; including working with the USAF Academy, which has courses and effects-based simulators. The USAF has no M&S career field designation like the Army does, but the presenter sees no need for a similar USAF designation. Instead, Mr TEMPALSKI is focused on building an M&S and experimentation certification course, which will include Technology, Entertainment, Design (TED) talk-like content.

Mr TEMPALSKI expressed frustration at bureaucracy being problematic, i.e., new contracts take 18 months to award. The presenter is working on mechanisms for building an M&S integration lab to integrate and verify and validate (V&V) mid-technology readiness level (TRL) technologies that can work around the contract delay issue.

Mr TEMPALSKI is working many objectives in support of NATO:

- Building a centralized repository of releasable models
- Getting more foreign national positions at the USAF Academy
- Starting to build courses at the USAF Academy that take less than 180 days because they qualify as a temporary duty station (TDY) rather than a change of duty station
- Upgrading the Warrior Prep Center (WPC) including updating effects-based simulators and rebuilding the WPC with higher resolution simulations
- Building a virtual document tagging system that automatically resolves security classification guidelines
- Opening up 2 seats for each NATO nation at the annual USAF M&S Summit. The 2024 Summit will focus on training and be held the first week of May in San Antonio, TX.

2.3 Session 1 – CD&E

Session 1 Chair – Dr Hans JENSE (Ministry of Defence, The Netherlands)

2.3.1 Paper #1 – M&S to Support Wargaming for Concept Development: Multi-domain Operations in Urban Environment Use Case; Dr Jan HODICKY, NATO Allied Command Transformation (NATO ACT)

Paper Abstract: *Development of a concept for Multi Domain Operations in the Urban Environment (MDO in UE) required validation effort to ensure that identified capabilities and the concept principles are pertinent to the future operating environment. In 2023 there was an ACT wargaming event bringing unique mixture of qualitative and quantitative assessment while seeking for a synergy between manual wargame and Modelling and Simulation (M&S).*

The MDO in UE concept is designed at the operational level; therefore, it requires inclusion of tactical activities in all five military domains while aggregating effects at operational level. The simulation framework was composed of a constructive simulation, an analytical tool harvesting M&S generated data, strategic dashboards using operational level metrics and a system dynamics model assessing the civil environment.

The paper commences with the Use Case description and M&S requirements. In later part, based on the design of M&S support to wargame and wargaming execution, the results of the validation event are portrayed. The last part describes lessons identified when looking for WG [war game] and M&S synergy in Concept Development arena.

Discussion: The focus of this paper is closely aligned with the theme of the Symposium. There is also a full report available. Dr HODICKY provided an overview of M&S in NATO Concept Development, which requires using different types of M&S and wargaming throughout the process. Figure 1 provides an overview of the uses of M&S across stages of NATO concept development.

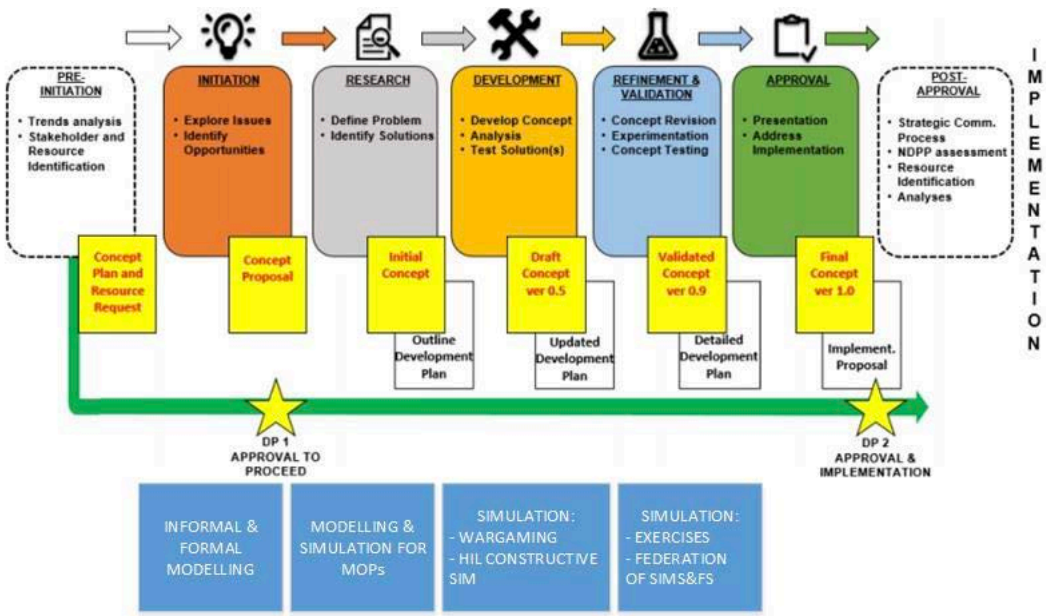

Figure 1. M&S in NATO Concept Development

The NATO Military Committee approved the Urbanization Capstone Concept in 2019 with the objectives of M&S support to the WG lifecycle and simulation to generate data for quantitative analysis to support the multi-domain operations (MDO) in the urban environment (UE) concept validation. The fundamental focus of this paper is answering the question, does the proposed concept enable NATO to be better prepared for MDO in UE?

Three teams participated in a 2023 Allied Command Transformation (ACT) wargaming event. Capability cards (CCs) were used to set expectations for participants. Each team developed their own courses of action (COAs) based on the same CCs. The CCs are selected by the planners by working backward from a desired end state of the scenario. Each capability card has a long list of effects. If the team leader wants to invoke the capability card, they have to follow the rules to achieve the desired kinetic and non-kinetic effect(s). Conceptual maps / mind maps were used for formulating simulation conceptual models because there isn't time to employ more formal processes and formats. Moreover, the team leaders aren't familiar with such processes and formats.

See the paper for key takeaways. One challenge is that the six main principles defined for each wargame and the metrics to assess achievement have to be identified manually.

Recommendation: See summary recommendation 1.

2.3.2 Paper #2 – Simulating the Impact of Navigation Waypoint Accuracy on Soldier Performance; Dr John GRAYBEAL, U.S. Army DEVCOM C5ISR Center

Paper Abstract: *The maturation of multiple technologies has made it possible to display various types of augmented reality (AR) information in Soldier night-vision devices (NVD). However, augmenting NVD optics with additional information will only improve Soldier performance if the information is sufficiently accurate. Further research is needed to better understand how Soldiers will perform when provided with realistically imperfect information. The U.S. Army DEVCOM C5ISR Center researched how Soldiers perform when NVD systems contain realistically imperfect navigation assistance. In such cases, Soldiers will need to leverage both their own navigational skills and system guidance to maximize their efficiency. In this experiment, 40 U.S. Army Soldiers navigated through a simulated urban environment using two desktop monitors. The primary monitor displayed a simulated NVD sensor view. Soldiers were instructed to follow an assigned route through the urban environment, which was displayed to Soldiers via a simulated paper map on the secondary computer monitor. Soldiers were instructed to navigate through the simulated environment as quickly as possible without deviating from the assigned route. They completed 48 scenarios with four different types of AR assistance. In the first experimental condition, Soldiers navigated without any AR assistance. In the second, AR navigation waypoints were perfectly placed along the route and were visible in the NVD sensor feed. In the final two conditions, Soldiers received AR navigation waypoints, but the waypoints were displayed with either minor or more severe geospatial registration accuracy errors. This enabled evaluation of the potential effects of providing Soldiers with realistically imperfect information, relative to unassisted Soldier performance and to Soldier performance with perfect AR guidance. The experiment's primary metrics of Soldier performance included the time to complete each route and deviations from the assigned route. All AR navigation assistance was beneficial to Soldiers, but the benefits were progressively reduced by greater errors in geospatial registration.*

Discussion: Dr GRAYBEAL presented an investigation of the operational effectiveness of application of augmented reality (AR) in soldier night-vision devices (NVD). The approach uses virtual prototyping—creation of a computer simulation of a product for presentation and/or evaluation. For addition of AR to NVD to be considered effective, Dr GRAYBEAL argued that the AR information presented to the soldier must be perceivable, interpretable, relevant, timely, and accurate. The presenter noted that AR can occasionally give bad or inaccurate information. Figure 2 shows the expected outcomes when using automated aids across a range of automated reliability.

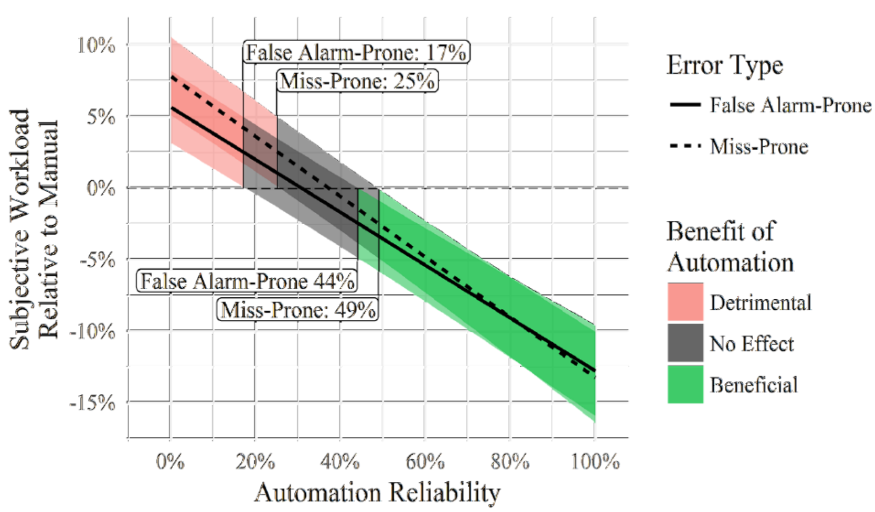

Figure 2. M&S in NATO Concept Development

Effectiveness in the use of AR in NVD was evaluated by comparing the ability to navigate a path under unaided (virtual navigation simulation) and aided (with simulated AR navigation overlay) conditions of varying quality/degradation. The experiment involved 4 groups of 10 soldiers (40 total subjects) through 48 simulated scenarios for a total of 1920 navigation paths. Experimental results confirmed expectations, where soldiers performed better with the augmented reality capabilities and when the augmented information met the criteria of perceivable, interpretable, relevant, timely, and accurate. Refer to the paper and presentation for full details of the measurements, analysis, and findings.

Dr GRAYBEAL identified limitations in this type of experiment that could affect generalizability of the findings. Such limitations include: differences existing between navigation skills in a virtual versus a real environment; limitations in using simulation, such as lacking other perceptual feedback (e.g., proprioception and tactile cues); being forced to stay on roads; and being tested in isolation with only a map and without any other navigational assistance they might have from teammates. Moreover, differences in stress or motivation may have affected how soldiers used the system. A more immersive simulation environment, perhaps using virtual reality and a 360°treadmill, might create a more effective environment for the navigation task.

Overall, this was a good example of the kinds of evaluation needed as new technology is introduced into existing capabilities. It must not be assumed that new technologies, including highly touted automated decision aids, will automatically provide better human performance without such careful assessments.

Recommendation: See summary recommendation 1.

2.3.3 Keynote Speech; VADM (ret) Ann E. RONDEAU, President, Naval Postgraduate School

VADM RONDEAU's presentation focused on her perspective relative to the Naval Postgraduate School (NPS). NPS is one of only two US military graduate schools, the other one being the Air Force Institute of Technology (AFIT). NPS has approximately 2,500 students. As an example of NPS' focus on technology, the USMC Commandant's focus on education established seven doctorate seats in technological fields of study. VADM RONDEAU is particularly concerned about "moving targets," by which the presenter means what we don't know yet. The presenter provided examples of recent, global surprises, including the COVID-19 pandemic and the biggest land war in Europe in 75 years where forces are running out of ammunition and employing kinetic force in addition to the new non-kinetic force. The presenter emphasized how narrative has become part of the battlefield in Ukraine and the Hamas-Israel conflict.

VADM RONDEAU charged the attendees as M&S practitioners to:

- Understand, analyse, and assess future challenges
- Use M&S to get us to what we don't know, need to imagine, and might miss
- Challenge ourselves and those we brief to understand our model assumptions, math, and analyses
- Be realistic about limitations and see them as opportunities for improvement
- Not limit our experimentation space because humans can be more variable than we might allow
- Get involved with the line operators and leadership to present better future representations with more "friction and heat" to improve preparation for operational contexts
- Use simulation to avoid catastrophic outcomes because the future is unconstrained

VADM RONDEAU closed by reminding us that data collection, science, and analysis are essential to understanding but so is teaming and that we need expert M&S leaders urgently. Our ethical purpose should be to understand facts and not false narratives, question assumptions, and ask whether the things we think we know from the past are still true today.

2.4 Session 2 – Training Effectiveness

Session 3 Chair – Mr Arjan LEMMERS (Royal Netherlands Aerospace Centre NLR)

2.4.1 Paper #4 – The OdySSEy to Transform Mission Readiness and Training Effectiveness; Mr Nicolas SMITH, BAE Systems

Paper Abstract: *Introduction/relevance to the Symposium: BAE Systems have developed a Single Synthetic Environment (SSE) which focusses on multi-domain, large scale training exercises that better prepare the armed forces for complex high-intensity warfare. We have developed the SSE to be easily configured for multiple scenarios, across the whole mission cycle.*

Rationale: Military simulation standards are at a crossroads. Project OdySSEy has evaluated the most complex issues, and challenges around simulation including fidelity and connectivity, creating a hybrid architecture that ensures we build upon an interoperable, open architecture, and scalable solution. To tackle these challenges, BAE Systems has leveraged technology from multiple small and medium enterprises (SMEs), operating in both the defence and non-defence markets.

Description of methods employed and results obtained: Project OdySSEy harnesses capabilities from over 8 different SME companies, including gaming technology, cloud computing, data analytics, biometrics, artificial intelligence and machine learning. These capabilities have been integrated and developed in an AGILE way, with sprints and iterative requirement updates, enabling rapid solution evolution.

Conclusion: Much has been learnt from Phase 1 of Project OdySSEy: the benefits of cloud computing and Artificial Intelligence/Machine Learning (AI/ML); the power of collaboration with SMEs; and the need for hybrid simulation standards. Phase 2 will build on this success, evolving the optimal hybrid architecture to support greater simulation technology ambitions.

Discussion: Mr SMITH presented the design and development of a Single Synthetic Environment (SSE) for use in collective training to address what was described as principle operational drivers, including: limitations on physical space; asset availability; distance (training with peers at distance to support multi-national, multi-level training); avoidance of monitoring training and tactics by adversaries; affordability/environmental impact; and the need to increase live training effectiveness by supplementing it with synthetic training at the needed level of fidelity, complexity, and security to deliver mission readiness. Project OdySSEy is intended to be a proof of principle, path-finding project dealing with principles of functional and developmental complexity, necessary fidelity to create a sense of immersion among participants, and security to lay a foundation for future development and use. Figure 3 gives an overview of the simulation design and an example screenshot of the developed user interface.

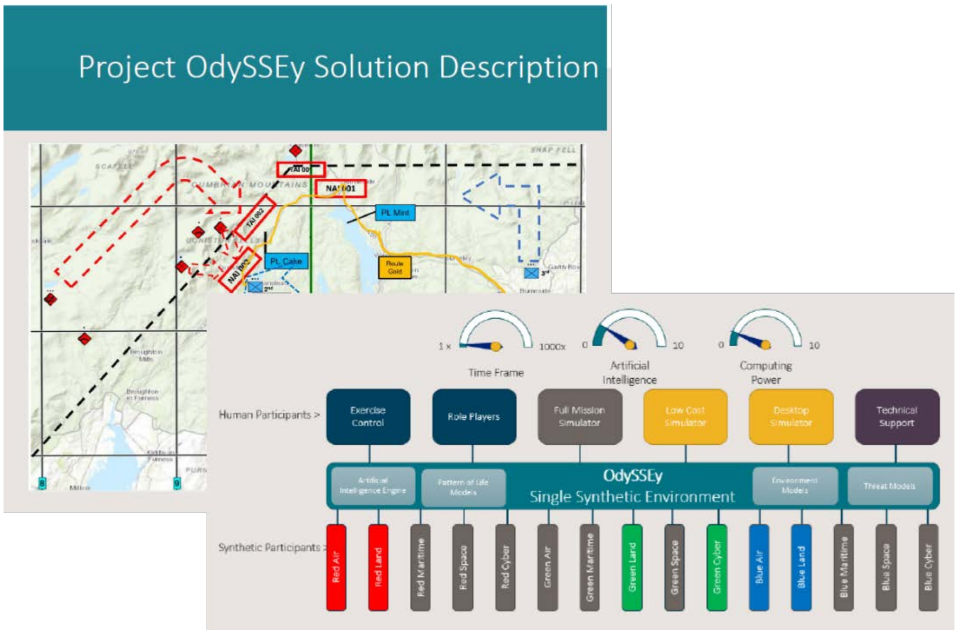

Figure 3 . OdySSEy Single Synthetic Environment Design Concept and Implementation

Beyond the integration of various simulation technologies, a key aspect of the project was to apply and evaluate innovative development practices, particularly with respect to the make-up of the team. The project brought together a diverse set of skills and expertise to create "cognitive diversity" with developers from several small and medium enterprises (SMEs) together with specialists in human factors, software development, and simulation. This diversity broadened collaboration and intra-team challenging of assumptions and common approaches to obtain novel and creative solutions. Development exercised agile processes, but with continual evaluation of the processes throughout. While meeting project objectives, a number of unexpected benefits were identified and explored, such as: training as a platform to accelerate technology readiness; competitive/gamified training; XR impact; augmented reality use cases; and pattern of life impact. Mr SMITH also openly shared limitations in the current capabilities, such as: entity visualisation limits; image generation visualisation discrepancies; simulator limitations; civilian behaviour stability; and data upload. Mr SMITH described numerous directions for future development and application of the simulation.

This paper and presentation gave an indication of what has become possible through application of modern software methodologies in an open-minded management culture and through current and emerging simulation and data analysis tools. Clearly, there are many opportunities for advancing M&S capabilities into broader warfighting domains through innovative open architectures and modern development practices. The ideas presented in this work complement those of Mr MCALINDEN's invited presentation in areas of technology integration and exploiting creativity of the workforce.

The paper identified use of simulation standards, such as the Distributed Interactive Simulation (DIS) and the High Level Architecture (HLA). The SSE seems to be focussed solely on integration in the simulation environment without considering broader interoperability issues, such as interactions with real-world command and control systems that have become an essential requirement for use of simulation in military training (as well as for analysis, experimentation, test and evaluation, etc.). For example, NATO has adopted the Command and Control Systems – Simulation Systems Interoperation (C2SIM) international standard (NATO Standardization Agreement 4856) for information interchange across C2 systems, simulation systems, and robotic and autonomous systems (RAS), while providing a foundation for information exchange across broader classes of systems. There is mention of control of unmanned systems, which also could benefit from consideration of the C2SIM standard.

Recommendation: See summary recommendations 1, 2, and 3.

2.4.2 Paper #5 – Increasing Training Efficiency through Automated Learner Performance Evaluation; Dr Robert SIEGFRIED, Aditerna

Paper Abstract: *We fight as we train. Providing warfighters with optimal training is therefore critical to succeeding within the 21st century military operating environment. However, current evaluation and analysis of training is not always optimal, mainly due to two factors: availability of highly trained subject matter experts as instructors and increasing data volumes generated from today's ever-evolving training environments that make comprehensive review and consistent trainee assessment cumbersome, if not impossible.*

We present a two-pronged approach with our training information management system RAPTOR that automatically computes trainee performance metrics to relieve instructors from routine tasks and to enable instructors to provide more immediate and consistent assessments.

While the current system already helps instructors to focus more on advanced evaluation and in-depth feedback, future system expansion would help groups of trainees to self-evaluate and discuss training events with their peers. Quantitative analyses of movement, communication flows and share of conversation within a simulated scenario provide valuable hints and can also work as conversation starters in the group evaluation of a shared training event. This paper presents results from working with US and European forces, utilizing the approach in training of fighter pilots and military medical personnel.

Discussion: Dr SIEGFRIED showed the correct perspective in the first sentence of the abstract: "We fight as we train." This is true only if we know that the training provides positive reinforcement of the needed skills and knowledge, which is often overlooked. In this paper and presentation, Dr SIEGFRIED emphasized the need for the means and metrics to evaluate training effectiveness. The presenter started by identifying challenges facing current training assessment, such as: increasing data volumes; increasing magnitude and complexity of reviews and assessments; insufficient automation creating dependency on availability of suitably qualified and experienced personnel; scarcity of expert personnel resulting in fewer training and assessment opportunities; varying qualifications and experience of instructors; and varying quality, scope, and documentation of training assessments. Dr SIEGFRIED argued that automated evaluation of training performance through advanced algorithms, analytics, and machine learning can minimize bias while enabling real-time feedback to optimize outcomes and resource allocation.

To address this challenge, Dr SIEGFRIED presented an improved feedback loop; see Figure 4. The presenter defined automated performance assessment (APA) as the use of advanced algorithms, analytics, and machine learning to objectively evaluate and measure the effectiveness of individual or group performance in training programs. Dr SIEGFRIED observed that team training performance is more difficult to address, having complex interrelationships among the members of the team resulting in difficult-to-measure synergies and weaknesses. The paper lists the main objectives of APA, such as standardization, providing augmentation to support instructors, enabling trainees to access training results, long-term tracking of performance, and immediate feedback, among others. As a step toward these objectives, Aditerna has designed the RAPTOR framework to ingest training data, calculate predetermined Measures of Performance, and provide scenario recommendations that ensure trainees master warfighting skills.

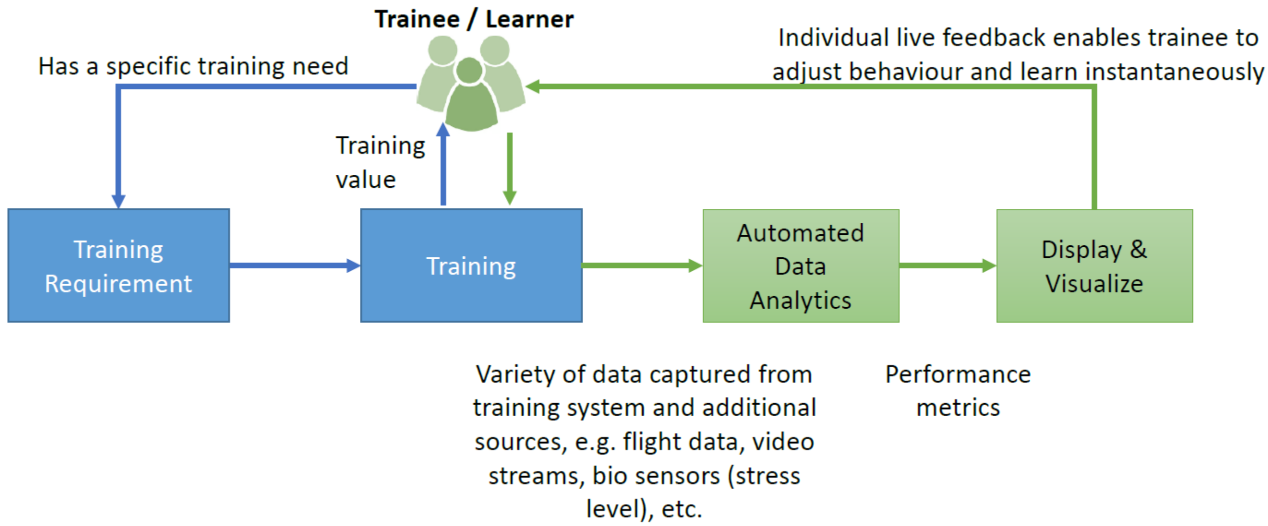

Figure 4. Interrelationships across Training Requirement, Trainee, Training, Automated Data Analytics, and Visualization to Improve Training Effectiveness

Future opportunities look to create opportunities for the individual learner through virtual reality, real-time assessment, adaptive training, integration of biosensors (e.g., to provide trainees with personalized, data-driven guidance from the sensors to help trainees improve performance and manage stress), and spatial analysis. For example, Figure 5 identifies a number of capabilities that can address the needs of APA.

- Generalized Intelligent Framework for Tutoring (GIFT)
- Air Force (USAF) Secure Live Virtual Constructive Advanced Training Environment (SLATE)
- Cervus XCaliber for marksmanship training
- Aptima Learning Analytics
- Aditerna RAPTOR

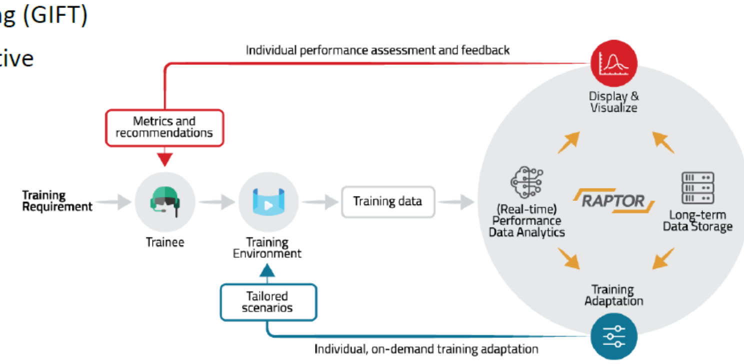

Figure 5. Available Technologies and Products for Integration with the RAPTOR Framework

The reviewers believe this paper and presentation emphasize the need for careful evaluation of the effectiveness of training environments. Too often there is an assumption that modernized simulation and computational tools will automatically (or, "automagically" as some see it) result in improved training. The reviewers see this as a dangerous assumption that puts warfighters at great risk. There must be careful, measured evaluation of capabilities to ensure negative training does not occur. Work presented by Dr SIEGFRIED is a good reminder that these are necessary efforts that may themselves be improved by greater automation, which also will need careful confirmation. As RAPTOR and associated tools evolve, it would be interesting to apply the approach to other activities briefed in this symposium, such as Dr GRAYBEAL's experimentation and OdySSEy capabilities, to determine its general applicability. The approach also raises important questions about the need for standards, such as standardized metrics for individual performance across military tasks (e.g., from established lists such as the US Army Universal Task List). As capabilities emerge for evaluating team performance, there is a growing need for that to include teams comprising human and robotic (and autonomous) systems. It raises the question, however, about whether the same techniques and metrics will apply.

Recommendation: See summary recommendation 1.

2.4.3 Paper #6 – Information and Data Protection in XR Training Scenarios: Straightforward Solutions to Comply with Modern Requirements; Mr Andrea D'ANGELO, Fondazione SAFE

Paper Abstract: *The use of Extended Reality (XR) in military training has been emerging for several years. Just like any innovative technology, this implies the use of lesser-known solutions, which bring additional security and privacy implications. The paper presents challenges and solutions identified during the development of XR simulations to support CBRN training, with a particular focus on information and data protection. Fondazione SAFE experience in chairing a NATO MSG-HFM-RTG on XR for CBRN and the internal development of XR CBRN training scenarios will be analysed.*

The paper presents key research questions addressed by SAFE in previous projects, to fully comply with NATO/ EU MS regulations on cyber security and data protection. An important takeaway is that, in training scenarios delivered via XR systems, the technical content (i.e., the characterisation and coding parameters of the CBRN event) could be considered classified and stand-alone headsets shall be prioritized when possible, adapting accordingly the coding of the XR experience. Classification shall also be foreseen for After-Action Report produced automatically by the XR training system to assess the trainees' procedures. Finally, the paper presents some of the challenges and solution to tackle in a simple yet effective way those issues and formulate valuable recommendations.

Discussion: Mr. D'ANGELO provided an overview of Fondazione SAFE (Security and Freedom for Europe) describing what the organization does, its activities with Extensible Reality (XR), and its Calvarina Testing and Training facility. The technical portion of the presentation provided a general security context, military use of XR and Virtual Reality (VR) technology, shortcomings identify in military applications, current cybersecurity guidelines for NATO, data and cybersecurity risks in designing a Chemical, Biological, Radiological, and Nuclear (CBRN) VR scenario, and solutions adopted by Fondazione SAFE.

Fondazione projects involving XR include: development and maintenance of rescEU CBRN mobile laboratories and rescEU CBRN detection, sampling, identification, and monitoring capabilities; realtime monitoring and sampling of CB (MoSaiC) threats for improved dynamic mapping, identification of vulnerabilities, and identification of response capacities; virtual enhanced reality for interoperable training of civilian and military CBRN operators (VERTIgO); NATO HFM-NMSG-354 Research Task Group on "Study, Design, Building, and Deployment of a CBRN XR Training Platform"; and a SAFE self-funded initiative for virtualization of the Calvarina facility, with creation of a CBRN VR training scenario.

Mr D'ANGELO pointed out that use of VR/XR technology is increasing in military forces for training (including maintenance of equipment) and under evaluation for use during combat operations. However, there are shortcomings in affordability, technological maturity, personnel, and cybersecurity. Regarding cybersecurity, there is a need for continuous assessment of XR/VR vulnerabilities. In NATO, Mr D'ANGELO expressed the need for dedicated technical standards dependent on the commitment of member and partner nations.

Basic data and cybersecurity designs for an application such as a CBRN VR scenario generally depend on use of wireless solutions, but these have increased risk of eavesdropping, interference, and unauthorized access. Software considerations include the presence of sensitive data and content regarding the scenario, use of appropriate encryption to access the library or launch the application, and third-party software monitoring to avoid unauthorised data transmission. Mr D'ANGELO noted that even when such a scenario does not contain classified data, the procedures undertaken by the trainees may be sensitive. Generation of automated after action reports (AARs) can involve data protection issues and aggregated data from multiple AARs could reveal gaps in trainee preparation and in defence capabilities.

Various networking approaches bring their own limitations: connection to the internet has the greatest weaknesses such as unauthorized/involuntary upload or transfer of information, data breaches, and the like; connection to an intranet is most vulnerable to insider attacks requiring robust security measures to be in place; and standalone offline execution allows the whole system and associated scenarios to operate in isolation from the external environment but limits availability of distributed operation.

Fondazione SAFE has applied a number of solutions to these challenges, including hardware protection through integrity of the headsets, password protection making the headset data-secure, and physical security of the headsets both inside and outside the headquarters. On the software side, techniques include a password prompt that is triggered when accessing the scenario; basic input/output system (BIOS) encryption, solid-state drive (SSD) encryption, and user password; generation of anonymous AARs to comply with data protection; and avoiding generation of classified aggregated data. Moreover, Fondazione SAFE employs trusted and heavily protected networks when necessary, offline standalone systems, and dedicated hardware to host all software related to the XR/VR solution to minimize network access from different devices.

In conclusion, Mr D'ANGELO emphasized the growing importance in establishing and applying appropriate cybersecurity practices in the development and application of XR/VR technology to create a cybersecurity culture at all levels. The reviewers can see that XR/VR/AR technologies certainly enable users to "go beyond the limitations of the real world" by immersing them in a mixture of "real" and "synthetic" worlds, enabling them to see the real world from new perspectives to gain new insights. This presentation emphasized the cybersecurity side of these technologies, an important step in focusing on "design-first" and "flash" later, an emphasis that is often hard to convince developers to take. Mr D'ANGELO pointed out an urgent need to define NATO technical standards, recommending that this be centralized to address requirements across a coalition.

Recommendation: See summary recommendations 1 and 4.

2.5 Session 3 – Decision Support - FMN

Session 3 Chair – LTC (ret) Elisseos MAVRATZOTIS (Pitch Technologies, Sweden)

2.5.1 Paper #7 – Advancing Modelling and Simulation in NATO Federated Mission Networking; Mr Kevin GALVIN, Thales, UK

Paper Abstract: *Federated Mission Networking (FMN) is a major NATO thrust to support interoperability, focusing on "Day Zero" interoperability in use/application of multi-national systems in coalition operations. Participants recognize the benefits of modelling and simulation (M&S) in supporting military operations, from training to rehearsal to decision support. The NATO Modelling and Simulation Group (NMSG) has supported the specification process for M&S-enabled capabilities in the FMN under previous MSG-193 and current MSG-201 activities over the past four years. Standards and best practices identified for integrating M&S into FMN include: (1) High Level Architecture (HLA), standardized under the Institute of Electrical and Electronic*

Engineers (IEEE) and approved under NATO Standardization Agreement (STANAG) 4603 Ed 03; (2) NATO Education and Training Network Federation Object Model (NETN-FOM); (3) Command and Control Systems - Simulation Systems Interoperation (C2SIM) (approved under STANAG 4856 Ed 01); (4) Modelling and Simulation as a Service (MSaaS). Developers and operators are employing these products in the current Coalition Warrior Interoperability Exercise (CWIX) testing the use of M&S in FMN. This paper presents an overview of each of the identified M&S standards applied to FMN development and considers their use in FMN implementation of Multi-Domain Operations (MDO).

Discussion: The paper was presented by Dr Mark PULLEN. FMN became essential with deployment of NATO International Security Assistance Force (ISAF) to Afghanistan. FMN isn't precisely a network but technologies and processes to interoperate using standards and uses spiral development. MSG-193 and MSG-201 designated an inter-WG Syndicate based on C2SIM, which integrated into spiral 5 with plans to expand technical capabilities in spiral 6. Integrating into FMN requires ready-to-run validated standards / practices, the proof of which is provided by participation in CWIX. C2SIM started with mission rehearsal rather than something more complex like collective training.

The operational requirements for FMN capabilities are specified in documents called Procedural Instructions (PI), describing what information is needed when, and by whom, in order to achieve a particular FMN function. The technical specifications supporting the actual implementation are part of documents called Service Instructions (SI). SIs for M&S define system interfaces based on standards, which for this effort includes:

- C2SIM
- HLA
- NETN FOM
- MSaaS

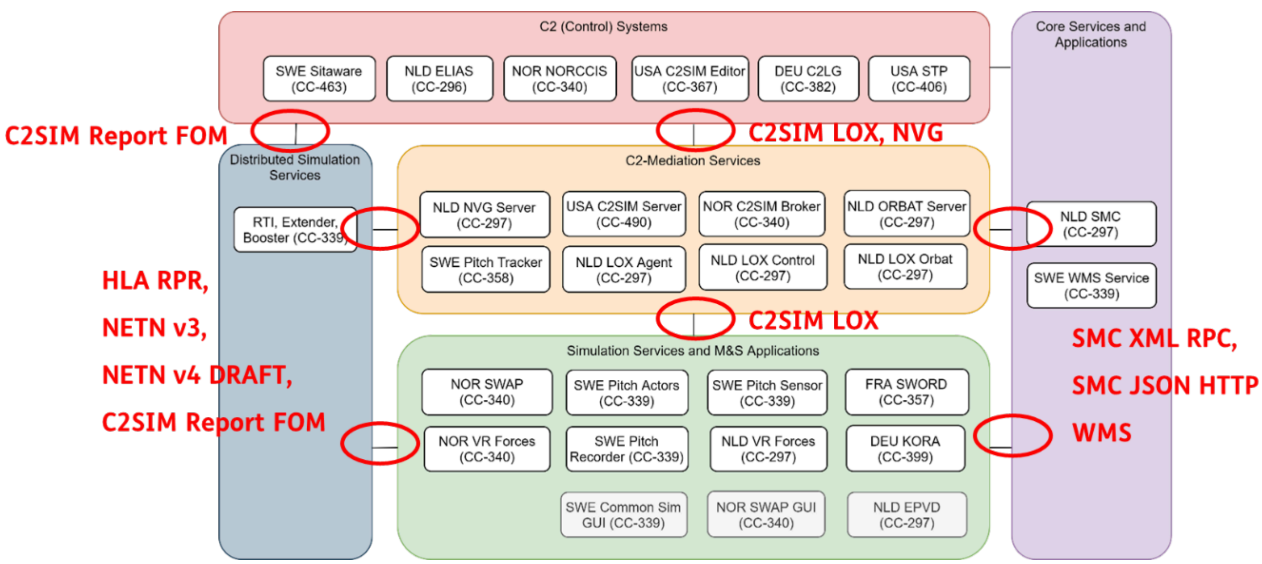

Figure 6. Overview of Tested Technical Interfaces

The technical interfaces, including standards-based ones, that were tested are illustrated in Figure 6.

NATO is committed to being MDO-enabled by 2030. More realistic mission rehearsal is the C2SIM goal for 2024. All the design and CWIX test case results are available in Tidepedia. The test results are automatically pushed to an analyst for validation. The team identified interoperability shortfalls including how simulation standards are implemented and shortfalls in simulation standards themselves. The latter are being fed back to the respective Product Development Groups (PDGs) and Product Support Groups (PSGs) in the Simulation Interoperability Standards Organization.

Recommendation: See summary recommendations 1 and 2.

2.5.2 Paper #14 – Using SysML Behavior Models in Simulation Environments for Mission Decision Support; Mr Shashank NARAYAN, Ansys Government Initiatives

Paper Abstract: *Systems Modeling Language (SysML) is an industry standard for defining system architectures. SysML-based architectural designs can also be used to model entire missions. These designs, however, need to be validated to ensure that they accurately reflect the complexity of the mission and its real-world operating environment, and ensure that mission objectives are met. An effective way to accomplish this is to execute the SysML models in a full physics-based simulation of that system's operating environment. In addition, SysML behavior models in a simulation environment can be used to establish and evaluate correlations between transitions in SysML workflows and those same events within realistic simulations. Collectively, these capabilities can be used to provide very effective decision support for entire missions, by predicting outcomes.*

Discussion: Mr NARAYAN led with the imperative for digital engineering, including the US DoD Digital Engineering Strategy. 30% of a program's lifecycle cost is in development; 70% is operation and maintenance. Better engineering decisions in development can reduce operations and maintenance; this is hampered by:

- Isolated models
- No common thread
- Duplication of effort

As illustrated in Figure 7, digital engineering requires:

- Open ecosystem
- Mission-centric
- Connected digital thread
- Multiple levels of resolution of models / simulations

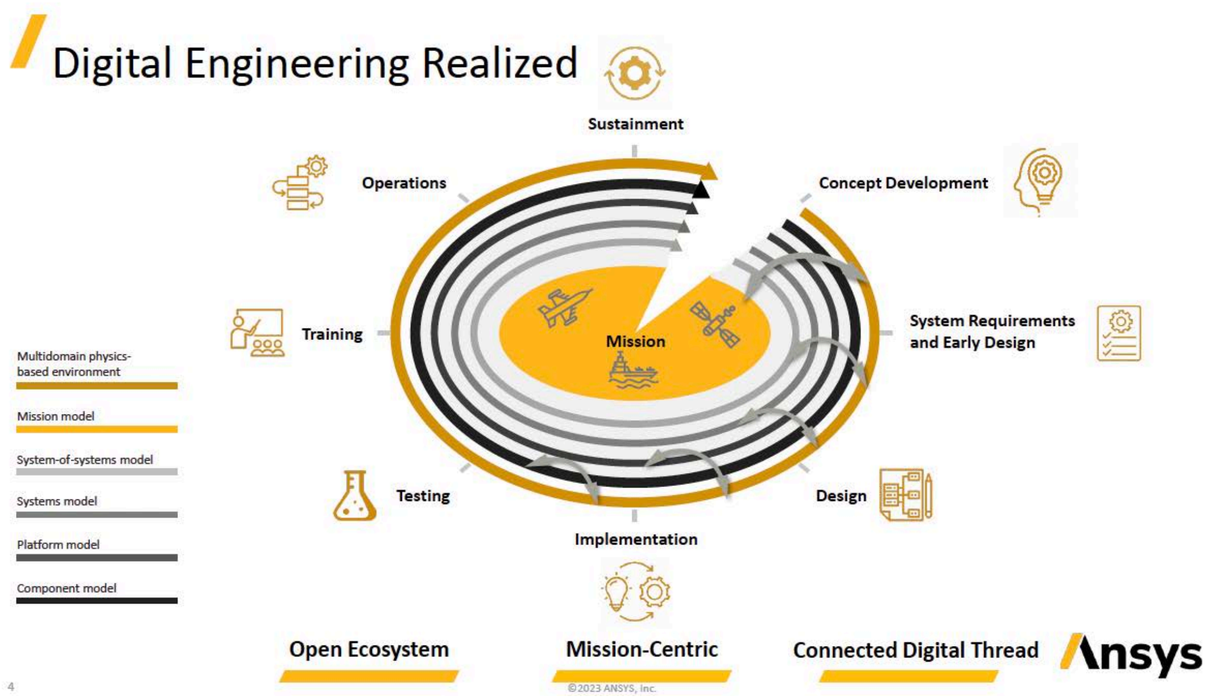

Figure 7. Digital Engineering Realized

It would have been helpful to highlight the triggers and guards in the paper. Sequence diagrams might have helped to serve as a conceptual model for the simulation. In response to a question about other simulation tools / frameworks that are easy to integrate with SysML, Mr NARAYAN cited their commercial product, Behavior Execution Engine, which has an API to connect to other simulation tools.

Recommendation: See summary recommendation 3.

2.5.3 Paper #9 – Ones and Zeros: Creating An AI-Driven, Data-centric Approach for All-Domain Operational-Level Simulations; LTC (ret) Matthew MARTIN, CAE USA Inc.

Paper Abstract: *As NATO and the member nations prepare for the highly-contested, near-peer potential conflicts of the future, the all-domain, data-intensive nature of these operations quickly outstrip the capabilities of legacy approaches to training, testing, and concept development. It is simply not possible to present sophisticated, multi-domain threat representations—with integrated effects from space, cyber, electronic warfare, and sophisticated red force tactics—on live-fly ranges or using legacy M&S tools. CAE has been developing a high-fidelity, all-domain, M&S environment, with red forces controlled by behavior graph-enabled generative AI, and tools to generate synthetically all the Command & Control Intelligence / Surveillance / Reconnaissance data, to enable training, testing, and concept development for near-peer, highly contested Battlespaces at scale. This capability has been tested at multiple large-force exercises. Initial results indicate that the process of multi-domain command and control, as well as intelligence exploitation, targeting, and planning can be improved in terms of speed and accuracy in executing the targeting process. This paper will review the development of these capabilities and the results of our testing. We propose the expansion of this capability to NATO operational C2 hubs such as Combined Air and Space Operations Centers (CAOCs) and Joint Task Force (JTF) components.*

Discussion: LTC (ret) MARTIN began with the challenge that intel operators lack data-centric training. Joint All-Domain Command and Control (JADC2) will need volumes of tailorable data for testing and to develop concepts and tactics, techniques, and procedures (TTPs). This type of training requires large simulations with many entities. Since 2016, CAE has been developing a data-centric virtual environment to enable training to the peer threat for C2ISR operators. They have tested and refined this capability in testing environments such as the NATO Coalition Warrior Interoperability Exercises (CWIX). The toolset, Virtual ISR Training Application (VISTA) claims to create essentially any C2ISR platform in the virtual space in any domain and attach sensors to generate data in NATO-standard formats. Joint All-Domain AI (JAD-AI) models the operator of the platform vs the platform itself. JAD-AI is illustrated in Figure 8.

VISTA plugs into computer-generated forces (CGFs), e.g., MACE, NGTS, AFSIM, and OneSAF, through its application programming interface (API). Using publicly available, unclassified data, a new model of a C2ISR platform can be built in a few hours. CGFs are responsible for implementing the sensor models, i.e., physics-based modelling. MACE has substantial capabilities for modelling operations in the electromagnetic (EM) spectrum.

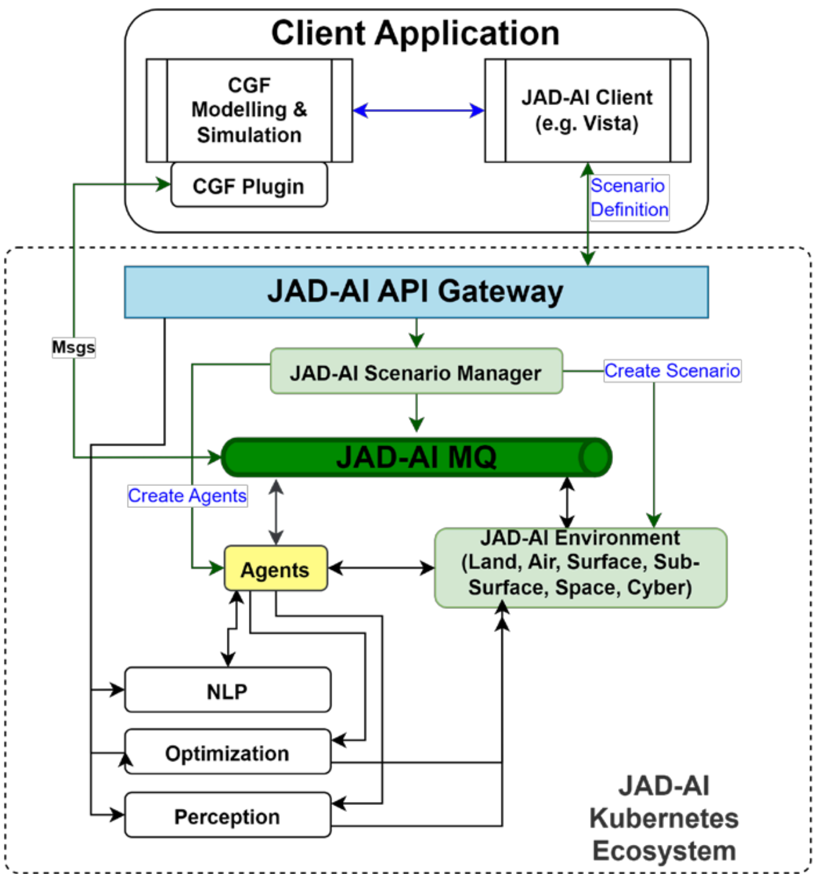

Figure 8. JAD-AI Architecture and Capabilities

There was insufficient time during the session to ask the following questions:

1. Are there any non-commercial participants in this effort?

2. *"A suitable battlespace Computer-Generated Forces (CGF) may already exist, and the GAI system will need to connect with it through an Application Programming Interface (API). The GAI communicates with the M&S system through analogous means. Open Standards make interfacing tasks relatively straight-forward for experienced teams once a set of common semantics between M&S/CGF, GAI and models is agreed upon."*

    a. Is the API defined?

    b. Aren't the common semantics a big problem?

3. *"It was also necessary to give them the ability to communicate with understandable language, including brevity codes or other shorthand."* What is the scope of the language allowed?

Recommendation: See summary recommendations 3 and 4.

2.6 Session 4 – Architectures and Interoperability

Session 4 Chair – Mr Lionel KHIMECHE (DGA, France, MS3 Chair)

2.6.1 Paper #10 – Leveraging Game Engines and Data Interchange Formats to Revolutionize Distributed Training; Mr Jerry HUGGINS, Ball Aerospace

Paper Abstract: *Commercial-off-the-shelf game engines, such as Unreal Engine and Unity, are rapidly expanding their capabilities and becoming an attractive option for modelling and simulation. Game engines have several benefits such as cost and licensing, first-class support for immersive display technologies, development pipelines geared around speed, and rich ecosystems. However, game engines lack physics fidelity, introduce coordinate system disagreements, and do not natively support military standard data interchange formats. Support for data interchange formats such as DIS and HLA is utmost if game engines are to be used in a distributed, networked training enterprise. While commercial data interchange plugins exist, they are expensive and difficult to modify and use. The Gaming Research Integration for Learning Laboratory® (GRILL®) developed a free and open source DIS plugin for both Unreal Engine and Unity which bridges the gap between game engines and the modelling and simulation enterprise. A rich framework was also developed whereby the plugin can be readily expanded to support additional standards such as HLA. Through the use of DIS, one of the primary drawbacks of utilizing game engines for modelling and simulation can be addressed. With help from the community, the plugin could become the industry standard, supporting many data formats.*

Discussion: Mr HUGGINS explained that GRILL®'s mission is to leverage commercial and Government off-the-shelf (GOTS) gaming hardware, software, and extended reality (XR) solutions to find solutions for USAF training and simulation needs with the following benefits:

- Game engines have a very low barrier to entry, i.e., they're free, but there are royalties and licenses for larger / more successful uses. These may change over the upcoming year.
    - Unreal may be going to a per seat license model.
    - Unity pricing might be based on a per download basis, i.e., downloads of the developed game / simulation.
- There's already a large labor pool, a marketplace, and open-source content for such components as:
    - Physics engine
    - Lighting / rendering engines
    - Audio / video engines

    - World builder tools
    - Developer tools
- Additionally, benefits include:
    - Community forum websites
    - High visual quality
    - XR support
    - Cross platform support for
        - Operating Systems (Oss)
        - Wearables
        - Handhelds
        - Consoles

On the other hand, the following deficits exist:

- Distribution method challenges in the military context
- Providing tech support within the military context, especially for classified projects
- Security concerns
    - New features and bug fixes
    - Unknown third-party developers
    - Plugins that use prohibited connectivity, e.g., Bluetooth
    - Getting an authority to operate in a classified environment may take 9 months to a year and bear added cost
- Geospatial resolution required for defence applications
    - But there are new game engine geospatial solutions, e.g., Cesium and ArcGIS
- 32 vs 64-bit precision, although Unity and Unreal are starting to support 64-bit
- Physics fidelity is not sufficient for some defence applications, so it may be necessary to use external high-fidelity tools that might not integrate well with the game engine
- Network communications
    - Data exchange issues can be mitigated by using DIS or HLA standards
    - GRILL® used NPS' OpenDIS and plans to develop an HLA plugin

Recommendation: Invite an updated presentation if / when they build an HLA plugin. See summary recommendation 4.

2.6.2 Paper #13 – Towards a Synthetic Environment Ecosystem for Test & Evaluation; Dr Robert SIEGFRIED, Aditerna

Paper Abstract: *As part of the future UK defence requirement for Test and Evaluation (T&E), Federated Synthetic Environments (FSEs) will be used to a greater extent. The Defence Science and Technology Laboratory (Dstl) envisions that this may be achieved through the creation of a Synthetic Environment (SE) ecosystem for T&E that is effective, promotes coherency and consistency, is readily sharable and reduces duplication of work.*

The aim is to provide multi-fidelity, multi-layered SEs that bring together the capabilities (layers) needed for a certain task at the required fidelity. Also, users must be enabled to conduct T&E in new ways or on emerging systems, such as increased use of digital twins, or T&E on Artificial Intelligence systems.

A broad industry team was tasked by Dstl to investigate this conceptual approach, a whole ecosystem of products, services and processes. This includes content produced by multiple suppliers, with novel arrangements to facilitate re-use, catalogues for accessing content, tooling to compose this content into FSEs, runtimes, deployment and hosting services, and analytical tools. Another key element is the ability to easily include content from multiple sources, including international partners. Beyond technical considerations, this requires proper governance and protection of intellectual property rights. Hence alignment with NATO standards and related NATO efforts, such as establishing a Modelling & Simulation (M&S) as a Service Ecosystem, is critical.

This paper summarizes the investigation's preliminary output covering: needs for such a SE ecosystem for T&E; information on relevant approaches and technologies and; outlines an architecture for a future SE ecosystem for T&E. Relations to NATO efforts are investigated and options discussed how the individual efforts could be aligned to avoid redundant effort and to benefit the envisioned ecosystems.

Discussion: Dr SIEGFRIED began by detailing the following concerns about defence T&E:

- Systems of systems (SOSs) spanning multiple providers and organizations
- Requirement for an ever-increasing range danger area
- Evaluating capabilities as complex and interconnected systems of systems (SOSs)
- Using operationally representative threats in a cluttered / contested environment
- Ability to collect, store, and analyse the large amount of complex data
- Enabling technology exploitation
- Implementing capability spiral development
- Evaluating innovative and new technologies

The briefing focused on phase 1 of the overall effort, the discovery phase. Phase 2, which just wrapped up, was refinement and initial implementation of approaches identified in phase 1. Future phases will evolve from completion of phase 2.

Phase 1 tasks:

- Review UK Ministry of Defence (MOD) policy and strategy docs
- Align basic terminology and concepts
- Identify specifics of T&E and especially how they're different from training
- Identify related national / international approaches
- Develop an architecture for an SE ecosystem for T&E (see Figure 9)
    - The architecture still needs to be aligned with MSaaS architectures.

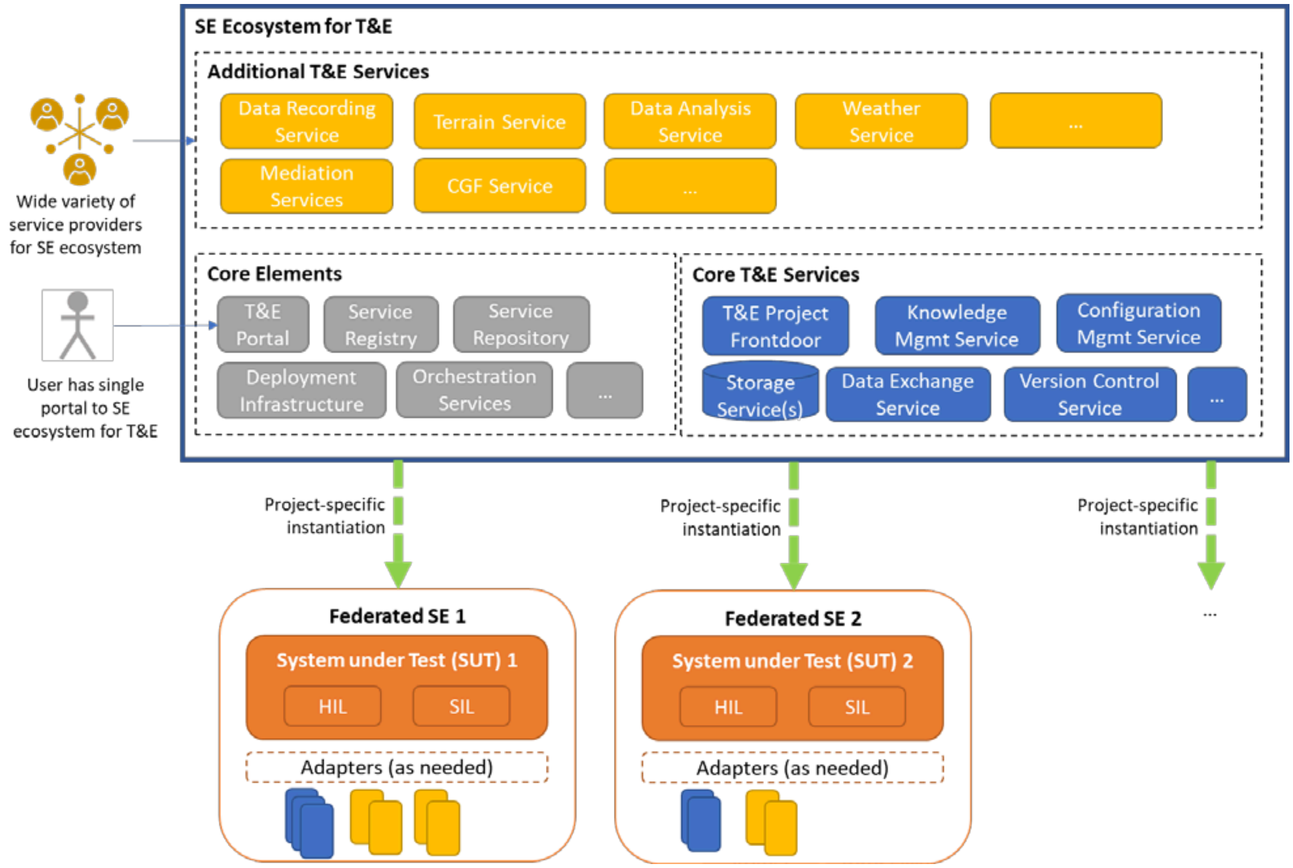

Figure 9. Architecture Outline for an SE ecosystem for T&E

Moving from singular federated synthetic environments (FSEs) to a synthetic environment (SE) ecosystem is a challenge. Demand for establishing an MSaaS ecosystem is growing. It's important to move from singular events to continuous T&E.

Recommendation: See summary recommendation 1: invite an updated presentation on phase 2 and future phase plans.

2.7 Session 5 – Decision Support

Session 4 Chair – COL (ret) John FERRELL (Lockheed Martin)

2.7.1 Paper #16 – Simulation of the Deployment of Forward Arming and Refuelling Points for Decision Support; Mr Philip MUURMANS, Netherlands Aerospace Centre - NLR / Swedish Defence University

Paper Abstract: *Forward Arming and Refuelling Points (FARPs) play a crucial role in supporting aircraft operations by facilitating deep penetration into enemy territory and ensuring sustained presence in the area of operation. The logistical units responsible for establishing FARPs face the challenge of timely deployment, since the necessary equipment must be on-site and operational before the aircraft arrive for resupplies. A crucial element in FARP operations is thus the selection of suitable locations, and taking into account the time*

that is needed to get there by truck. M&S enables the exploration of different options during the planning phase of FARP operations.

The contributions of this paper are threefold. First an initial model of the relationships between the factors surrounding the success of FARP operations is presented. Second, a simulator prototype is constructed based on the model, consisting of Monte Carlo simulation to quantify associated risks and discrete event simulation to simulate the time aspects of such operations. The simulator is implemented as a Python-based simulation program that interprets real-world map images, simulates FARP operations, and generates data such as helicopter fuel consumption and FARP survival probability. Third, the use of the simulator for decision support regarding the deployment of FARPs is demonstrated in two use cases.

Discussion: Mr MUURMANS described a simulation developed specifically for analysing deployment of FARPs to best support an operation. FARPs need to be located in positions that optimize refueling of aircraft performing missions in an operations area, both to facilitate deep penetration into enemy territory and to ensure sustained presence. Figure 10 shows relationships between successful FARP operations and overall operational effectiveness. The paper and presentation show the factors and lower-level operations that contribute to or detract from achieving FARP success. The simulation is used to perform an analysis seeking to find feasible FARP locations based on the nature of the terrain at the location; that is, determining where reasonable helicopter landing zones can be found. Given those locations, the simulation then conducts a mission planning and execution phase to evaluate the location's ability to support flights from the forward operating base (FOB) to the area of operations, then to a FARP, and then back to the FOB. The conditions are further modified by running non-dispersed FARPs (a single FARP location) and dispersed FARPs (three FARP locations are used such that helicopters have to choose which FARP to use based on availability). Simulation outputs include the amount of fuel used by each helicopter, the amount supplied by each FARP, and survivability of each FARP. Experimentation considered two use cases: (1) trade-off between fuel consumption and FARP survivability, where FARP survivability is determined by the survivability of the transportation of fuel overland to the FARP; and (2) fuel use in dispersed versus non-dispersed FARP operations.

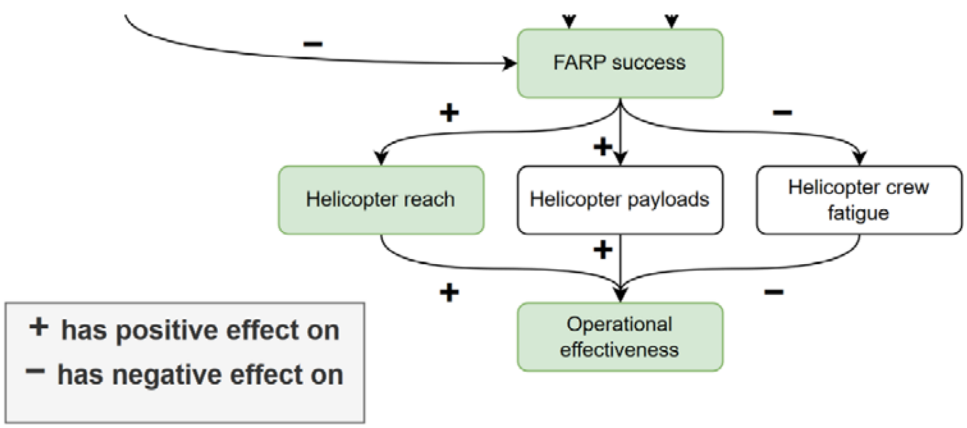

Figure 10. Impact of Successful FARP Operation on Operational Effectiveness

The paper and presentation provide detailed plots and analysis of the results from runs of the simulation for these variants. Mr MUURMANS concluded that simulation was useful in examining FARP operations from these differing perspectives and show that a simplified model can have benefit to logistics planners. In future work, the researchers plan to add additional considerations, such as: bringing in resupply of ammunition, human factors, and different helicopter and mission types; providing more realistic mission environments using up-to-date input data;

integration of realistic flight models; and adding a geographical information system (GIS) application to enhance the location analysis portion of the work. The model is designed to enable consideration of varying demands (e.g., from routine sustainment to resupply during active combat). Enhancing logistical planning and execution through simulation execution and analysis is a critical combat multiplier in today's world of complex multi-domain warfare.

Recommendation: See summary recommendation 1.

2.7.2 Paper #17 – AI-enabled Logistics Intelligent Decision Support (A-LIDS); Mr Damian GALLEGOS, Lockheed Martin

Paper Abstract: *The goal of AI-enabled Logistics Decision Support (A-LIDS) is to enable operational commanders with informed logistics decision making for command and control of crewed and autonomous resupply operations. A-LIDS leverages AI models, trained on simulation generated data lakes and future fleet wide data streams, to enhance the commander's ability to observe and orient to both the existing situation and future battlefield conditions. A-LIDS relies on statistical techniques and AI services for decision support that:*

- *Provide overall readiness measures across all classes of supply*
- *Tracks and predicts key measures of effectiveness*
- *Provides LogScores: a comparative evaluation of logistics courses of action*

Training effective AI for A-LIDS relies on data from many warfighting situations and platforms. To capture the logistics posture, data must be both broad and deep, capturing not just individual unit actions, but the entire logistics chain. To model the future fight, we train the AI using synthetic data from a U.S. Army validated simulation. This simulation exceeds the limitations of data quality not currently available in the real world. The A-LIDS dashboards provide improved situational awareness through recommendations, predictions, and optimizations. The intent is to stimulate autonomy, reduce cognitive load of commanders, and improve mission success.

Discussion: As motivation for this paper and presentation, Mr GALLEGOS provided historic examples of logistical challenges that affected military operations, such as the Battle of Monterrey and D-Day, both dealing with logistical support to conduct of operations in enemy territory. Such examples become pertinent in our modern era when operational concepts include emplacing and sustaining forces within enemy engagement zones. Logistics has always been critical to military operations; contested logistics presents the most extreme challenges. Moreover, it is projected that there will be significant employment of autonomous systems for logistical delivery in those contested environments by 2040. Mr GALLEGOS went on to argue the current JADC2 lacks sufficient logistics situational awareness to enable effective planning and execution. Decision support to enable such awareness requires relevant data, and a source of such data can be from simulation—using representations of the battlespace and the battlespace dynamics to generate information on demand under differing situations.

In this case, Lockheed Martin is using the US Army WARSIM model to generate data from accelerated fires scenarios varying joint logistics platforms, vignettes, missions, and outcomes for training AI algorithms. The AI-enabled Logistics Intelligent Decision Support (A-LIDS) system is expected to enhance sustainment decision support to accelerate joint sustainment operations. Given sufficient fidelity in the synthetic environment, the same algorithms could ingest real-world data in real-time to near real-time to support ongoing evaluation and continuous planning during an actual operation. However, the reviewers believe that proving the resiliency of the AI/ML models to provide equally effective support in actual operations compared to that obtained from synthetic data remains to be done. The work expects to leverage AI to predict readiness across classes of logistics, provide course of action evaluation with explainability, provide actionable insights, and suggest efficiencies. The researchers have prototyped a dashboard with various information visualization and analysis to examine operational utility. Components of the dashboard include such capabilities as asset readiness prediction, parts survivability analysis, and synthetic data generation. Refer to the paper and presentation for a description of the overall architecture and examples of representative data in the prototype dashboard.

Mr GALLEGOS concluded by inviting others to collaborate with them in this endeavour. The presenter is particularly interested in obtaining feedback from the U.S. services and subject-matter experts in the dashboard design. The reviewers agree that advancing logistics decision support is an important, data-rich area for application of AI/ML technologies.

Recommendation: See summary recommendations 1 and 3.

3.0 DAY 2 (FRIDAY, 20 OCTOBER 2023)

3.1 Invited Presentation: Natural language AI for Military Decision Support and Swarm Control for Autonomous UAS Trained in a Combat Simulation; Mr Daniel KALLFASS, AIRBUS Defence and Space GmbH

Abstract: *The future of warfare is undergoing transformative changes through the integration of AI-assisted command systems and unmanned technologies, which will have a significant impact on combat operations and the required speed of military decision-making cycles. Future decision-support systems will assist military decision-makers in evaluating threats, developing optimal courses of action for their forces, and even executing actions through collaborative swarm behaviors of autonomous systems. To enable these systems, the combination of modeling & simulation, and advanced Deep Reinforcement Learning (RL) techniques will play a crucial role.*

This paper presents the results of several studies conducted by the German Army Concepts and Capabilities Development Centre and Airbus. These studies evaluated the adaptation and utilization of simulation and AI techniques to train an AI agent capable of acting as a battalion commander in an Army combat or controlling a swarm of UAVs in an ISR mission using the RL-optimized simulation "ReLeGSim". The AI agent generates natural language commands using a language model to execute actions within ReLeGSim, enhancing communication between human advisors and AI systems while incorporating objectives and doctrines into the AI reasoning process. Through a military doctrine-aware feedback function, the AI agent assesses and improves its behavior during each training cycle.

Once trained, the AI agent can be applied to real-world scenarios, developing courses of action alternatives to a battalion commander derived from the learned AI agent policy, or directly executing them in autonomous systems to control a swarm of UAVs. This research serves as a foundation for equipping AI agents with the ability to uphold military doctrines and rules in future operations.

Discussion: Mr KALLFASS opened by reviewing ground-breaking deep reinforcement learning events. The team used DeepMind’s AlphaStar for use cases that exhibit the following challenges: fog of war, effect of actions that can be far in the future, many unpredictable random effects, no real training data, and limited computation power. Figure 11 illustrates the ReLeGSim architecture.

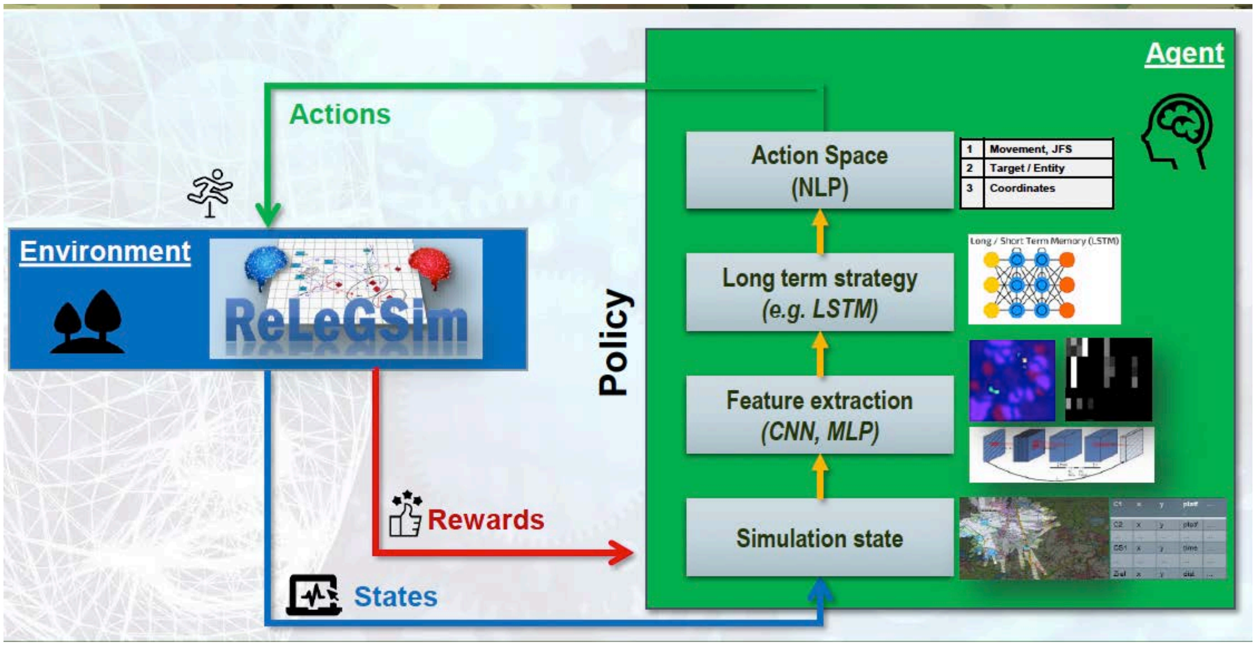

Figure 11.Deep RL Overview & Architecture

The ReLeGSIM natural language processor (NLP) has 100 different words and maximum 8-word-long sentences to task the AI. The performance of the model and the goodness of the reward function are closely related. The best reward during the study consisted of a combination of rewards: sentence reward, teaching the objective.

The decision support use-case had a vision of super-human strategies to speed up the observe-orient-decide-act (OODA) loop. They developed analysis dashboards (see Figure 12) to visualize the agents' behaviour for the purpose of determining achievement of this goal.

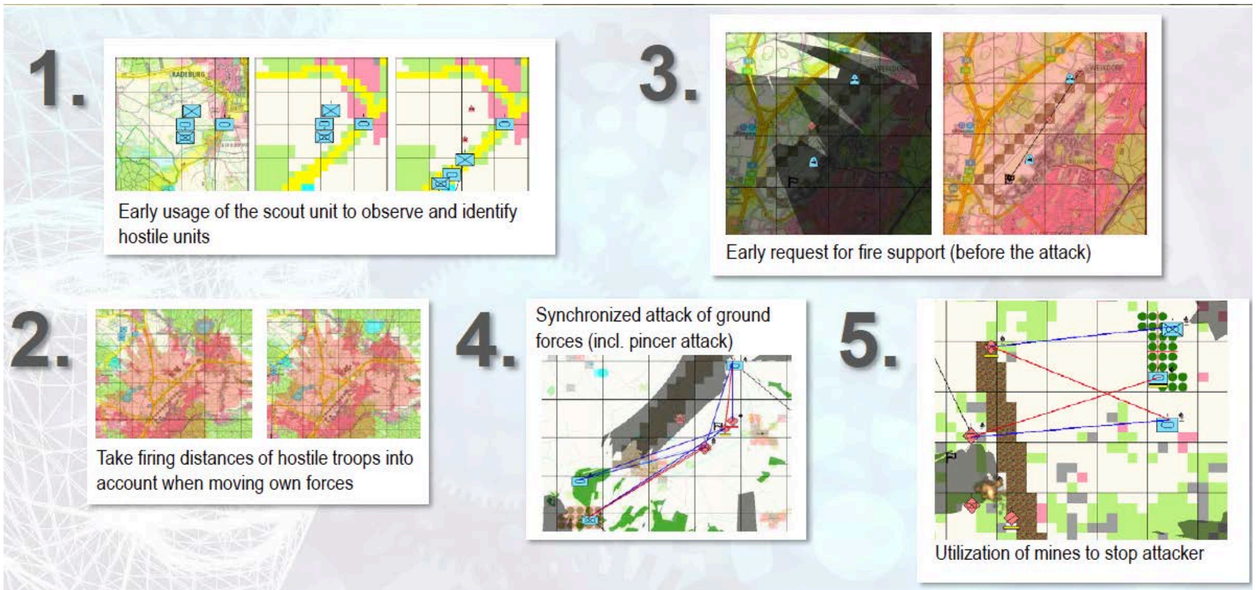

Figure 12. Examples of Observed Behaviours

The UAV swarm autonomy use case tested whether autonomous UAVs stay within range, follow orders, and perform self-maintenance, i.e., returning for new batteries. The team demonstrated that natural language can be combined with RL and used as a human-on-the-loop interface. There is still the issue of explainability of such large neural networks.

Recommendation: See summary recommendation 3. The Way Ahead section of this paper could provide valuable input to the summary recommendation 3 tasking.

3.2 Session 6 – Adapting M&S for National needs

Session 6 Chair – COL Miro COLIC (Vice Dean Military Education, Croatian Defence Academy)

3.2.1 Paper #18 - Optimizing Smaller Western Countries' Resource Allocation in Remote Warfare via System Dynamics Modeling; MAJ Cedric CRANINX, Naval Postgraduate School – Belgian Army SOF

Paper Abstract: *Remote warfare consists of specific components interacting to maximize operational effectiveness. Many smaller Western countries opt for this model to counter threats from a distance, reduce risk, and curb financial costs. However, due to their strategic culture and limited resources, these countries cannot fully employ the model, which reduces their operational effectiveness. To determine how small Western countries can maximize their use of resources, in this paper, system dynamics modeling, and simulation is used to analyze the impact some remote warfare components have in a counterinsurgency. Remote warfare occurs in substate conflicts where Western countries support local partners acting as counterinsurgents. The end state of counterinsurgency military operations is to defeat the insurgents militarily. Consequently, the study uses the insurgents' size as the key variable measuring the operational effectiveness of two types of support: training support and intelligence support to a local partner. Equations quantitatively and qualitatively model insurgents with an information advantage competing against counterinsurgents with a size advantage. Data from the Islamic State insurgency case is used to validate the model's fit over a simulated 36-month run and draw conclusions. The study finds that intelligence support is highly operationally effective, while training support by itself is not.*

Discussion: MAJ CRANINX motivated this presentation by explaining that remote warfare refers to a country's military being removed from the contact line versus the sole use of remote technologies (remote control war). What form of support is primarily responsible for making the remote warfare system effective? The measure of effectiveness is size of insurgent force. The insurgent force is assumed to have the information advantage, but the counter-insurgency (COIN) force typically has the advantage of force size. The model uses system dynamics modelling together with simulation. This work analysed the following scenarios:

1. No support – slow increase in COIN find & fix capability over the simulated 36-month period, which is insufficient

2. Training support – increasing COIN size is similar to scenario 1, but still insufficient

3. Intelligence support – increasing levels of intelligence support more significantly increases the COIN find & fix capabilities, actually tripling, with level 3 resulting in a decrease of insurgent force size

4. Training and intelligence support – this shows slight improvement over scenario 3

Information remains a function of COIN size and insurgent force size, but it is more a function of the added synergy of multi-int capabilities provided by the remote warfare intelligence support. Without external intervention in the insurgent conflict, the insurgent force's information advantage counterbalances their size disadvantage.

In the reviewers' opinion, the paper itself is very long and detailed. It could have benefited from synthesis. It also draws a conclusion about "gaining control of the political environment and addressing the underlying social and political issues that gave birth to the insurgency in the first place" unsupported by the reported research. But it is an argument for integrating this simulation with an human, social, culture, behavior (HSCB) simulation to explore the interaction of these processes; for example, refer to the earlier Peace Support Operations Model (PSOM). This paper and presentation are also good examples of work attempting to avoid the "COIN hangover" brought up in Mr MCALINDEN's invited presentation on the first day of this symposium.

Recommendation: See summary recommendation 4.

3.2.2 Paper #19 - Desktop & Cloud MH-60R Helicopter Training Simulator, Designed for Rapid Modification for Different Nations; Dr Robert RICHARDS, Stottler Henke Associates, Inc. (SHAI)

Paper Abstract: *The US and other Navies in conjunction with industry has developed and deployed a training tool used for training for the US and the international versions of the MH-60R helicopters. Since the MH60R helicopter can be configured differently for different nations, the training also needs to be adapted to the specific versions of the helicopter. Nations flying or soon to be flying the MH-60R include Australia, Denmark, Greece, Spain, India, and South Korea. The tool, called the Operator Machine Interface Assistant (OMIA), is primarily an expandable, easily modifiable low-cost PC-hosted desktop or cloud-based crew. OMIA provides much of the cockpit interface for the front seats and the Sensor Operator station, providing training in most aspects of helicopter operations except flying, this includes but is not limited to navigation operations, radio operations, emergency operations, RADAR, ISAR, ESM, FLIR, and both active and passive acoustics.*

OMIA is designed to be flexible, allowing for low-cost, rapid modification since the software and crew interfaces evolve, and need to be adjusted for different nations. OMIA's flexibility has allowed SHAI to rapidly make the modifications necessary to keep the various non-US versions of OMIA consistent with the respective non-US versions of the MH-60R.

Discussion: Dr RICHARDS presented a low-cost, modifiable training tool for the MH-60R helicopter. Modifiability is needed so the training can be readily adapted to national variants of the MH-60R, as well as to accommodate evolution of the crew interfaces over time. The capabilities are provided through the Operator Machine Interface Assistant (OMIA) emulating many of the non-flying tasks needed in the aircraft, such as the navigation, radio operations, emergency operations, and sensor operations. The form factor facilitates anytime, anywhere training. OMIA complements training conducted through other means, such as classroom, computer-based training, aircraft simulators, and time in the actual helicopter. Dr RICHARDS described reducing overall training costs from reduced need for simulator and actual helicopter time, as well as increased overall effectiveness from the availability of anytime, anywhere training via laptop, cloud, touchscreens, VR, and other readily available devices. The paper and presentation provide numerous images of actual cockpit and device interfaces for comparison with screenshots of OMIA emulated interfaces.

Although there has been considerable emphasis on measuring training effectiveness in other presented papers in this symposium, the reviewers noted that such discussion was lacking in this paper and presentation. For example, it would be interesting to see results of study of initial knowledge acquisition with and without OMIA in the overall training sequence (from classroom to real-world aircraft), as well as knowledge retention if there are periods of time when crew members are not able to man the actual aircraft. Additionally, there may be an interesting cross-over in application of the automated performance assessment techniques presented by Dr SIEGFRIED (paper #5) to evaluate training effectiveness of OMIA.

Recommendation: See summary recommendation 1. The tool architecture and capabilities provide an example of low-cost, modifiable interface design that may be applicable to other aircraft and may be a suitable object for training effectiveness studies using techniques presented in this symposium.

3.3 Session 7 – Digital Twins

Session 7 Chair – Mr Simon SKINNER (Thales, UK)

3.3.1 Paper #20 - Digital Twin: a Bridge between Simulation and Real World in the Maritime Environment; Dr Giovanni Luca MAGLIONE, NATO STO Centre for Maritime Research and Experimentation (CMRE)

Paper Abstract: *This publication describes the CMRE effort on the development of the Digital Twin (DT) architecture for Maritime Unmanned Systems, with the goal for the DT to serve as a bridge between modelling and simulation and the real world applied to MUS technologies and to explore how cyber-physical bridging is achieved during the execution of two NATO exercises.*

The maritime domain, especially underwater, is a complex, "non-human-friendly" domain. MUS applications face challenging environmental conditions. Nonetheless, there is an increasing interest in MUS from the operational community; MUS can perform dangerous missions without putting people in harm's way. New MUS capabilities are developed and these systems are becoming more advanced and autonomous, being able to take over tasks normally performed by expensive manned platforms. Industry, together with research institutes and academia are developing new MUS (prototypes) in a rapid pace. This combination makes MUS a demanding but also a very relevant research topic in support of NATO exercises and operations. The Digital Twins concept provides a cost-effective means to analyse the operational capability and interoperability of existing MUS while at the same time reducing the risks involved with actual deployment and use of MUS. Digital Twins further facilitates the development of requirements for future MUS by providing the opportunity to easily change or upgrade MUS capabilities using modelling and simulation. For the above reasons DT is a very active and evolving research area for CMRE, and also an area that has high interest and visibility within the technical and operational community.

CMRE developed a DT prototype and deployed it during two NATO exercises REPMUS22 and DYMS22. These exercises took place in Portugal, September 2022. This DT prototype, which is currently further being developed and expanded, serves as the first DT framework prototype for the nations involved. An important development within the DT is the implementation of the Collaborative Autonomy Tasking Layer (CATL) interoperability protocol. CATL (emerging in the standardization agreement STANAG 4817) makes the DT system seamlessly interoperable with existing Command and Control (C2) systems on-board ships.

Discussion: Dr. MAGLIONE described work to investigate the use of a digital twin (DT) for test, evaluation, verification, and validation of Maritime Unmanned Systems (MUS), particularly with respect to underwater autonomous systems. The presenter pointed out the growing need in NATO for maritime situation awareness. The underwater domain presents particular challenges, such as different, and arguably more complex, environmental conditions than surface or air operations, limited communications, and the need for remote sensing. The paper lists a number of benefits of a MUS DT for the technical and operational communities, including: use as a situational awareness tool to improve insights in execution and operation; investigating effectiveness of deployment alternatives and tasking; assessing interoperability between different MUS in a larger scenario; simulating communication network performance to analyse the configuration of different MUS for different tasks; and others.

The development effort used the Distributed Simulation Engineering and Execution Process (DSEEP; IEEE 1730-2022) and the NATO Architecture Framework (NAF) in a spiral and agile process. The team completed a prototype in October 2022 and added improvements through 2023. Figure 13 shows components of the simulation environment, both from the initial work and additions in 2023, using the HLA runtime infrastructure (RTI). The M&S federation included an "interoperability federate" serving as a bridge to provide message exchange with external non-HLA systems, such as C2. The foundation for this interchange is the C3MRE infrastructure and the Collaborative Autonomy Tasking Layer (CATL).

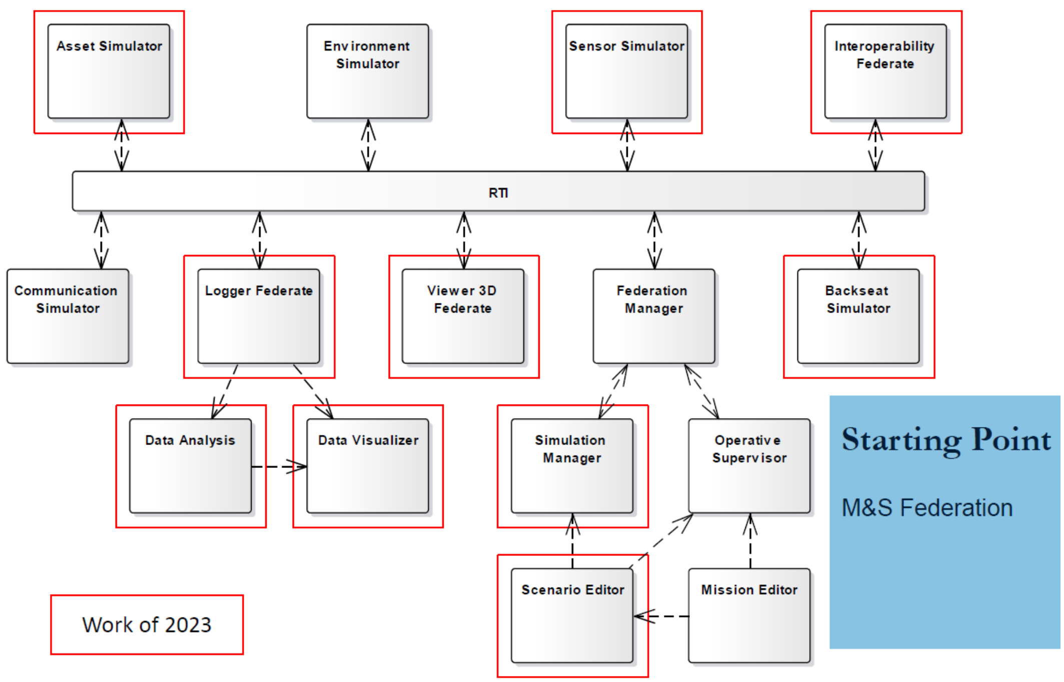

Figure 13. Components of the Simulation Federation for MUS Test, Evaluation, Verification, and Validation

The research team defined and investigated scenarios with the following objectives: clarify the operational picture; compensate for data incompleteness; CD&E; training; health monitor systems (HMS); and test and evaluation to verify and validate (TEV&V). The first and last of these are described in the presentation. One use case focussed on C2 operational awareness using 3D replicas of all platforms with emphasis on the data interchange needed across systems and platforms. Another use case focussed on testing, evaluation, verification, and validation using raw data collected during real-world operations to test systems and to verify and validate new AI modules before deploying to sea. The work visualized real assets in a 3D environment, augmenting real assets with simulated assets executing parallel missions and augmenting real assets with simulated sensors to evaluate detection performance.

In the area of data interchange between C2 and the simulation, the effort looked at status information (telemetry from and to C2), task administration (e.g., mine counter-measure survey and identification missions from the C2 system), and dynamic update of contacts and detections from and to the C2, as well as a shared world model (e.g., regions, missions) between C2 and the unmanned system.

The reviewers believe it is becoming increasingly clear that simulation is a valuable tool for evaluating systems, but there is a significant trade-off to consider. On the one hand, the real world presents too few opportunities for live testing, resulting in a small sample space under very limited real-world conditions. On the other hand, simulation can create a wide variety of test conditions and situations but lacks real-world detail (simulation is always a simplification) creating uncertainty in the validity and robustness of the test results. These trade-offs need to be carefully analysed and balanced to create as effective a test/evaluation regime as possible.

Moreover, the paper states:

> *A NATO working group belonging to Systems Concepts and Integration (SCI) entitled 'Enabling Federated, Collaborative Autonomy' is working to build up a common language between heterogeneous assets to create a system of task generation and exchange. This system has a dual purpose. Firstly, the system connects heterogeneous assets independent of the actual command and control in use. Secondly, the system orchestrates the data flow between the assets as efficiently as possible, taking into account the limited bandwidth available in the underwater domain." (p 6)*
>
> *"For this capability, A CATL Bridge was implemented that allows any CATL and C3MRE compliant asset to exchange information with the M&S federation, hence allowing the federation to move towards the implementation of the DT concept." (p 10)*

In this area, the work provides a basis for a parallel activity to examine the application of the C2SIM international standard (NATO STANAG 4856) addressing information exchange across C2 systems, simulation systems, and robotic and autonomous systems (RAS), while providing a foundation for interchange across broader classes of systems (e.g., cybersecurity systems, combat systems, sensor systems, etc.).

Overall, this is significant work reaching into challenges of multi-domain operations with the added complexity of integrating across live, virtual, and constructive systems.

Recommendation: See summary recommendations 1 and 2. Begin collaboration with other NATO groups who are employing C2SIM in the Federated Mission Networking (FMN) initiative to inform each other's work and to potentially reduce duplication of effort.

3.3.2 Paper #21 – A Concept for the Modelling and Simulation of Complex Urban Environments; Mr Matthew SMITH, DSTL, UK

Paper Abstract: *The urban environment provides challenges for military forces such as complex physical terrain; a population of significant size and density; multi-dimensional infrastructure and transportation systems; and complex communication and information systems. To prepare future forces for these complex environments, it is critical that they are accurately represented in future* *training systems, due to limitations in live training ("the Real World") such as scale, complexity and cost.*

The Defence Science & Technology Laboratory (Dstl) has researched methods of understanding, modelling and simulating the complexity of the urban environment to increase the effectiveness of the next generation of military training systems. Analysis of existing models and frameworks against anticipated training requirements has concluded that a single simulation approach is impractical.

A proposed solution is a 4-tier urban simulation architecture that combines virtual simulation, with constructive simulation to model Political, Military, Economic, Social, Information, Infrastructure, physical environment, and Time (PMESII-PT) effects. Each tier operates at a different level of geographical abstraction and temporal activity, allowing a more efficient use of computing resources.

This paper describes the core architecture, information flows, and efforts towards building a technical demonstrator to support UK military training and force development activities.

Discussion: Mr. SMITH described the challenges of modelling the complex urban environment, where there is an interplay of varying levels of abstraction across geography and time addressing the intermix of social systems, physical systems, and information systems. The work was informed by urban training considerations, as shown in Figure 14, along dimensions of the interconnected landscape model, ATrainP3 training categories, and types of urban operation (note: HADR – Humanitarian Assistance / Disaster Relief; COIN – counter-insurgency).

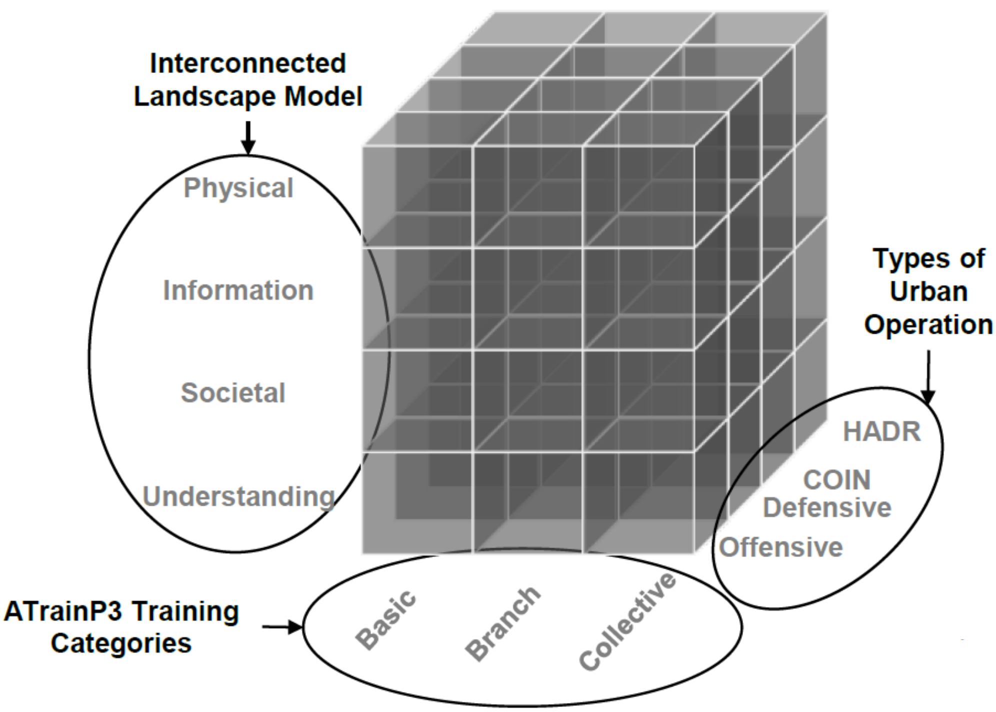

Figure 14. Urban Training Considerations (note: HADR – Humanitarian Assistance / Disaster Relief; COIN – Counter-Insurgency)

Researchers at Dstl have conceptualized the environment in four tiers: (1) an inner tactical zone focussed on detailed, dismounted, individual and small unit collective training operating in real-time with detail to the level of, for example, building interiors, deformable terrain, water, electricity, and gas services, and the resident population; (2) an outer tactical zone covering a larger area at lower fidelity, also running in real time, with consideration of ingress/egress routes, building exteriors with limited representation of interiors (e.g., floors, partitions), critical infrastructure (where actions taken in the outer zone can affect conditions in the inner zone), and populations represented as crowds; (3) a city region zone with wider political context involving all PMESII-PT effects, including disruption to infrastructure networks, and major population effects, and where large urban areas are split into multiple districts, longer-than-real-time actions are applied, and the population is represented as aggregate objects; and (4) a global information zone as the outermost layer connecting different city-scale model information environments and actions that influence such effects as sentiment and global diaspora.

To support live training for close-quarter battle, small-scale villages/towns are represented, but it is recognized that at this scale, it is not possible to represent the full complexity of urban warfare. Rather, there would be a limited representation of critical infrastructure, social, and information environments. For such breadth of requirements, the scale and complexity possible with M&S are essential. Today's modern cities are impractical to model in fine geographic detail but instead are divided into smaller geographic blocks.

Regarding the temporal abstraction, units of measure from minutes to weeks/months are selected appropriate to the functional tier. Considering the urban M&S spectrum, Mr. SMITH observed that no single training use case requires all tiers operating simultaneously.

Mr. SMITH described future work as consisting of two demonstrations with different focuses. They will move toward a 2025 concept of how to exploit the urban M&S ecosystem within defence operations.

The reviewers believe conceptualization of different layers and scales appears very appropriate for the problem and, furthermore, creates opportunities for multiple domains of warfare to interact, such as a cyber-attack bringing down the power grid, thus affecting the population pattern of life and disrupting military actions. Dstl is exploiting a long history of significant modelling and simulation in the PMESII-PT space, such as the earlier Peace Support Operations Model (PSOM). A major trade-off can be considered, from interconnecting existing or emerging models of the different layers or creating new capabilities from scratch (or some combination of the two).

Recommendation: See summary recommendations 1 and 4.

3.3.3 Paper #15 – Decision Support within Complex Urban Operations; COL Dr Peter HOFER, Theresa Military Academy Institute of Advanced Officer Training

Paper Abstract: *A commander's decision-making for operations in an urban environment requires support by specialized expertise and visualization assets to make best use of the process. This paper discusses the benefits of integrating decision-preparation and decision-making in the reality-virtuality continuum with special consideration of virtual reality and shows possible time savings.*

Discussion: COL HOFER described the problem with training for urban warfare versus the reality of urban warfare as training sites are too small and simplistic. The presenter described urban environments as being characterized by size, complexity, and "Triple-S": supersurface (e.g., multi-story buildings), surface (e.g., streets, sidewalks), and subsurface (e.g., subways, tunnels). Figure 15 illustrates these aspects of the urban environment. Contributors to complexity include violent opponents, difficult environmental conditions, extensive infrastructure on all levels of movement, the problem that opponents and features are mostly hidden from view, and the practical problem of being able to come up with detailed maps across the multiple levels and layers.

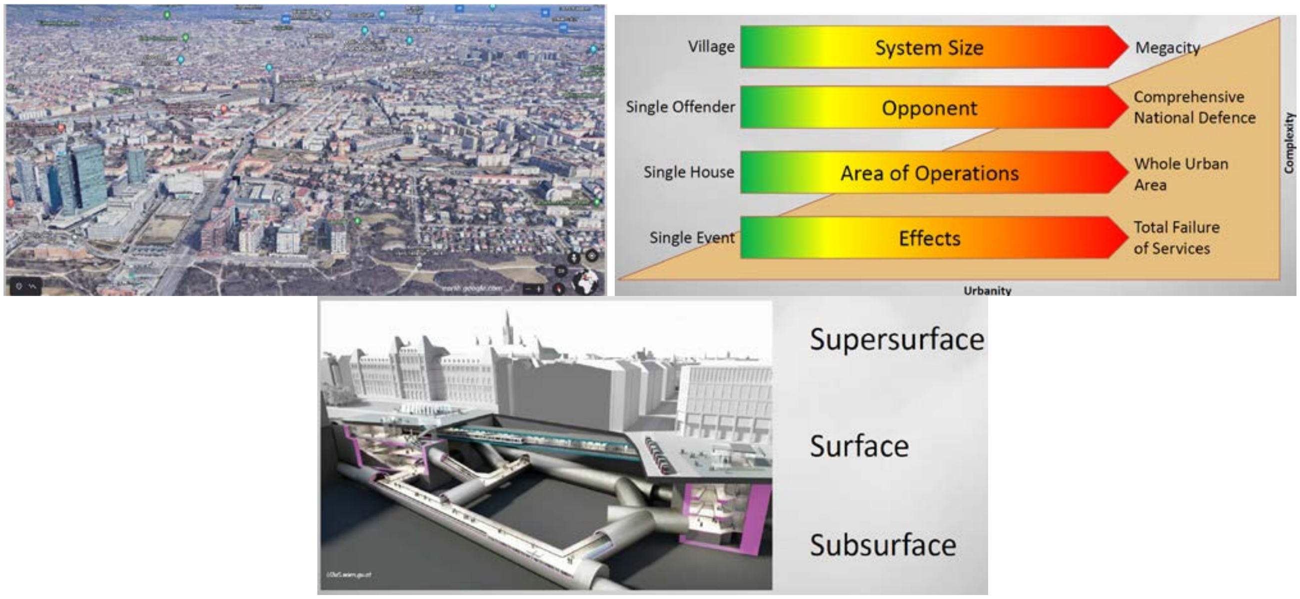

Figure 15. Key Characteristics of the Urban Environment: Size, Complexity, "Triple-S" (Supersurface, Surface, Subsurface)

COL Hofer observed that decision time can be improved by better visualization and collaboration in a virtual environment, when dealing in time-constrained and unconstrained conditions. Such visualizations and collaboration can be maximized by seamless interaction. The presenter described some preliminary experimental findings to support these statements. To move toward such a capability, COL HOFER described the Rapid Data Integration and Visualization (RApid) program but admitted that access to data is the key challenge to overcome. RApid is providing some innovative ways to visualize the urban environment as a basis for experimental evaluation. The main tool for enabling planning in an urban environment is the Triple-S Operations Mission Tool (S3OMT). If considered just for training, the presenter believes the capability does not have to be fully accurate but must provide enough realism to make soldiers aware of critical considerations. Eventually, such capabilities could support actual operations if sufficiently complete and accurate information about the urban area can be obtained and integrated.

As a basis for evaluation, COL HOFER described the cognitive load theory of Sweller, Ayres, and Kalyuga (Springer, 2011) based on *intrinsic load* (complexity of information) + *extraneous load* (info representation) + *germane load* (resources required for problem solving). The goal of digital interfaces would be to reduce the extraneous load while maximizing the germane load and reducing the complexity of the intrinsic load. They have begun experiments into use of VR for decision-making, which is showing a benefit, but this has not been evaluated conclusively at this time. COL HOFER concluded with the expectation that visualization will improve the quality of the perception of urban environments and reduce the time required to comprehend the situation in that environment. The presenter expects there to be improvements in content production that will reduce time for preparation of the digital model. While the presenter expects decision-making quality to improve, the decision-making time should not be shortened. The saved time can be used for other activities, such as gaining deeper insights for even greater gains in decision quality and resulting operational effectiveness.

The reviewers consider this work to have great potential and applaud the focus on measuring the benefits obtained from the technical approach. In an operational planning context, the reviewers agree that obtaining accurate information and maintaining currency of that information (e.g., due to deformation as a result of weapons effects) will be a significant challenge. Even so, use of the capabilities in training offers opportunities to improve warfighter insights into the operational situation and may prove transferable to an actual operation. That will need careful study and confirmation. Clearly, a workable strategy may be to start simple (e.g., a 50% solution) and add complexity over time, verifying continued warfighter competency throughout the process.

Recommendation: See summary recommendation 1.

3.4 Session 8 – AI

Session 4 Chair – Prof Andrzej NAJGEBAUER (Military University of Technology, Warsaw, Poland)

3.4.1 Paper #23 - Scaling Artificial Intelligence for Digital Wargaming in Support of Decision-Making; LTC Scotty BLACK, Naval Postgraduate School / U.S. Marine Corps

Paper Abstract: *In this unprecedented era of technology-driven transformation, it becomes more critical than ever that we aggressively invest in developing robust artificial intelligence (AI) for wargaming in support of decision making. By advancing AI-enabled systems and pairing these with human judgment, we will be able to enhance all-domain awareness, improve the speed and quality of our decision cycles, offer recommendations for novel courses of action, and more rapidly counter our adversary's actions. It therefore becomes imperative that we accelerate the development of AI to help us better address the complexity of modern challenges and dilemmas that currently requires human intelligence and, if possible, attempt to surpass human intelligence—not to replace humans, but to augment and better inform human decision-making at machine speed. Although deep reinforcement learning continues to show promising results in intelligent agent behavior development for the long-horizon, complex tasks typically found in combat modeling and simulation, further research is needed to enable the scaling of AI to deal with these intricate and expansive state-spaces characteristic of wargaming for either concept development, education, or analysis. To help address this challenge, in our research, we are developing and implementing a hierarchical reinforcement learning framework that includes a multi-model approach and dimension-invariant observation abstraction.*

Discussion: LTC BLACK expressed a vision of a superhuman wargaming agent that serves as the foundation for creating modern decision aid tools that can provide decision makers more accuracy, speed, and agility over the more traditional tools. This audience has considerable experience with wargaming and more than a little understanding of AI / ML, so the introductory content wasn't necessary and might have been reduced in preference to new information. The Atlatl combat simulation environment and OpenAI Gym Environment were used for this research as illustrated by the architecture in Figure 16. A video of the AI achieving the optimal solution was presented.

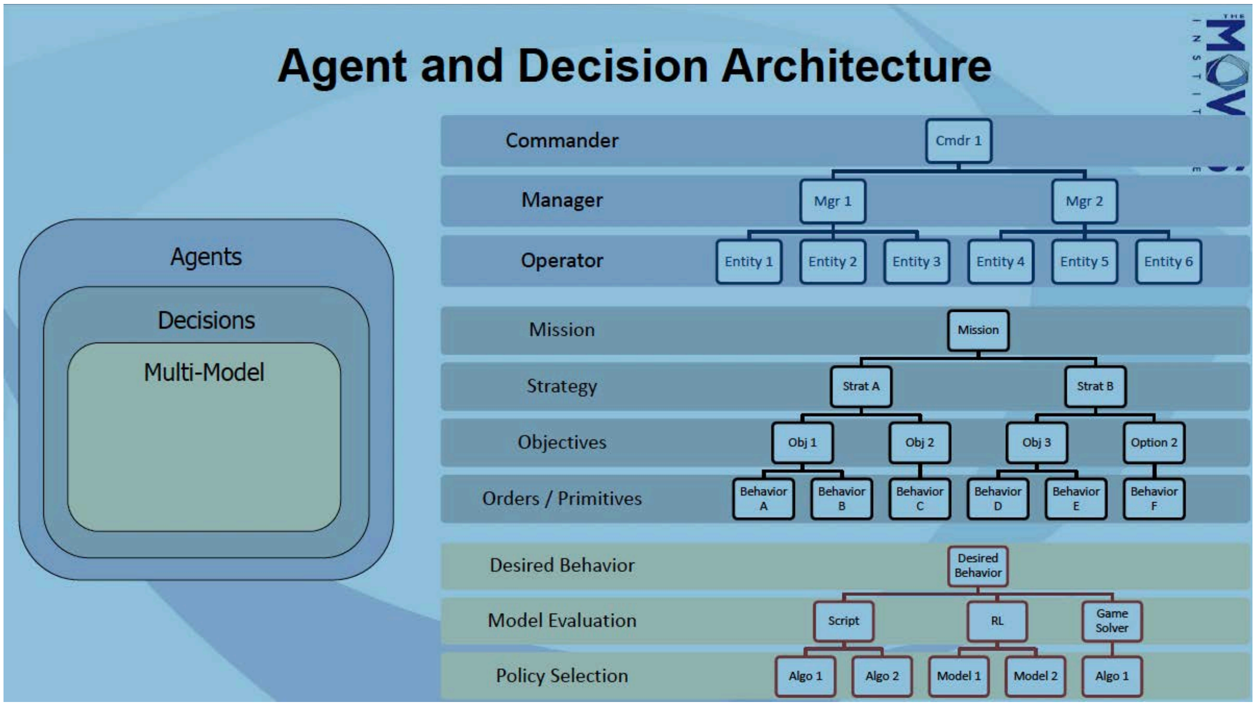

Figure 16. Agent and Decision Architecture

Hierarchical reinforcement learning (HRL) that allows for training multiple hierarchical layers of agents, decisions, and policies using different levels of observation abstractions was applied in the training framework as illustrated in Figure 17.

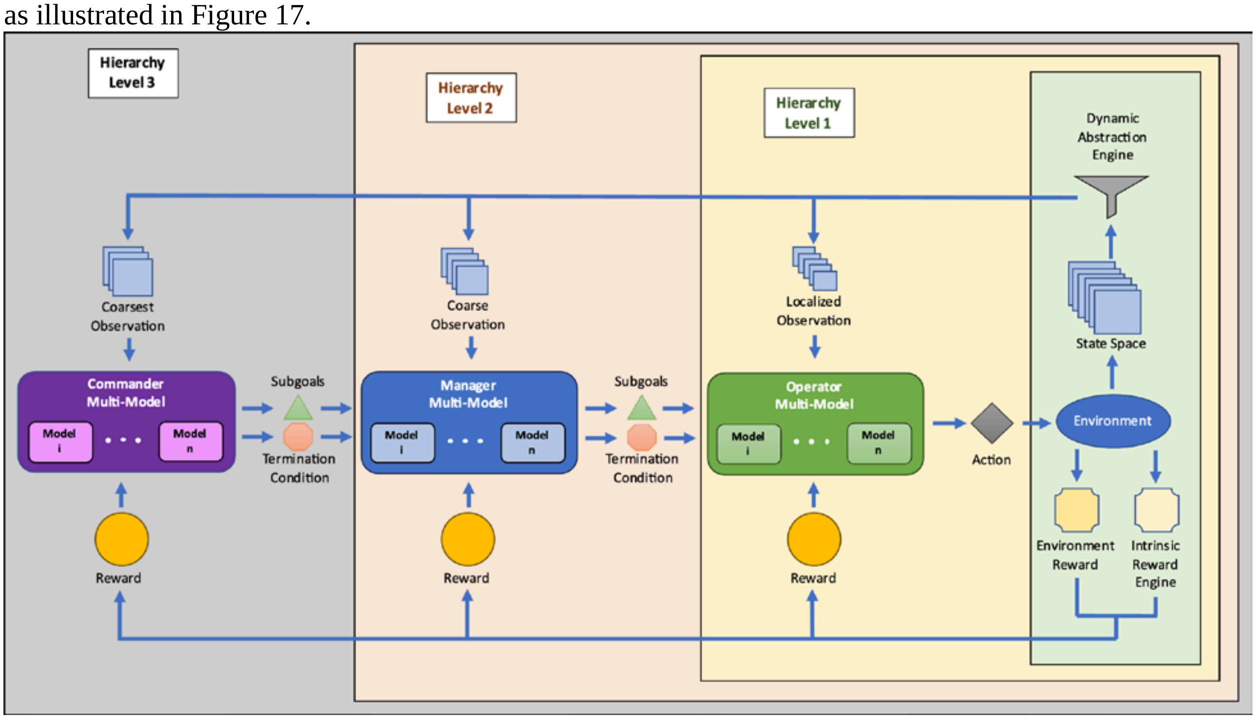

Figure 17. HRL Training Framework

LTC Black reported that the research to date shows promise with this approach to scale the application of AI to more complex domains such as combat M&S in support of wargaming. The multi-model framework has drastically improved the performance of the agents well beyond current state-of-the-art scripted or RL-trained agents within combat simulations. The presenter also indicated that individual specialized agents should be implemented with different RL techniques.

Recommendation: Focus on the details of the specific research and results. See summary recommendations 1 and 3.

3.4.2 Paper #24 - Multi-Dimensional Data Farming: Extending Data Farming for Multi-Scale Decision Support by Integrating Novel AI Techniques; LTC Stephan SEICHTER, Bundeswehr Office for Defence Planning, DE

Paper Abstract: *Traditional Data Farming (DF) consists of a toolbox of established analysis techniques that are available for an analyst-led study of a particular military operation in support of a single decision-maker. Multi-Dimensional Data Farming (MDDF) is a new and automated analytical process that provides accelerated support to military decision-making at multiple scales. At the strategic level, MDDF can inform decision-makers in planning long timescale campaigns, while at the tactical level, MDDF allows investigation of emerging technologies in shorter timescale operations. More importantly, MDDF explicitly addresses the interplay between a long timescale campaign and embedded short timescale operations, which is rarely tackled in the literature. MDDF extends DF by integrating novel AI techniques (Automated Machine Learning, eXplainable AI) and eXtended Reality visualization in an AI agent which automatically investigates the multi-dimensional parameter landscape and efficiently provides decision-makers with insight into the best, worst and most promising Courses of Action. We illustrate our new MDDF approach through a hybrid warfare scenario consisting of a Border Operation (interdiction of illegal migrants) embedded within a multi-faction (Blue, Red and Green forces) hybrid war campaign. Combining AI techniques exploring operations at multiple scales (domain, level, time) and boosting strategic and tactical understanding, MDDF innovates multi-scale decision-making.*

Discussion: LTC SEICHTER explained that Multi-Dimensional Data Farming (MDDF) automates simulation decision-making by the integration of innovative AI techniques and allows improved and faster decisions in highly complex multi-scale, multi-domain, and multi-level hybrid war campaigns. Data farming was codified by MSG-088, Data Farming in Support of NATO. MSG-197 is investigating a scenario where green supports blue against red using the strategic level, multi-faction, multi-domain combat model Attrition, Cyber, Epidemic (ACE) and the tactical level multi-agent-based simulation of the border operation built in Map Aware Non-uniform Automata (MANA) as illustrated in Figure 18.

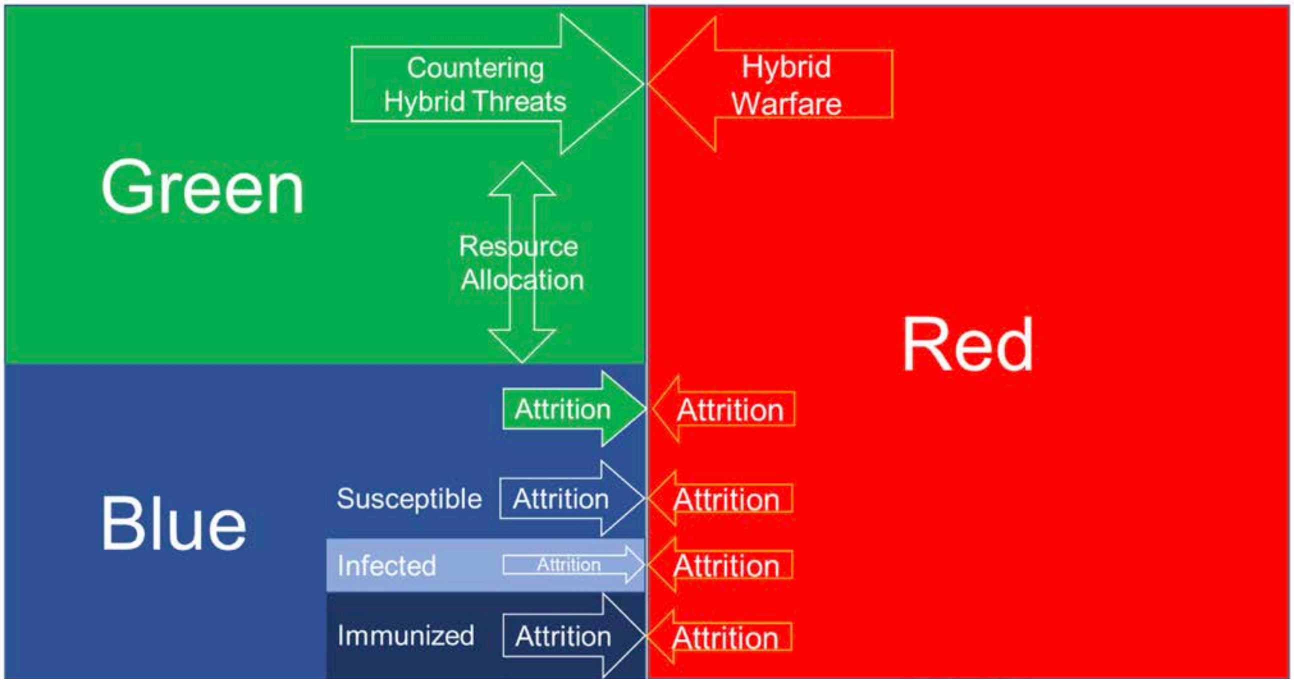

Figure 18. Unconventional Campaign Scenario

Classical data farming with data sets the size that are output by such a simulation is challenging, so a multi-model approach is indicated, in this case using explainable AI, automated machine learning, and Bayesian optimization with Gaussian processes.

Design, Analyze & Visualize Experiments (DAVE) performs design of experiments, analyses the optimal balance of forces between the unconventional campaign and the border operation, and visualizes the temporal playback of the unconventional campaign with spatio-temporal playback of the border operation to gain insight. DAVE is based on data farming services from MSG-155. It's currently a standalone tool but will be extended to a demonstrator for CWIX 2024. More operations will be added in the next phase to increase its applicability.

Recommendation: See summary recommendations 1 and 3.

3.4.3 Paper #25 - Empowering Military Decision Support through the Synergy of AI and Simulation; Dr Joost VAN OIJEN, Royal Netherlands Aerospace Centre

Paper Abstract: *In military tactical operations, there is an increased demand for machine-based decision support capabilities to support human decision-makers. Artificial Intelligence (AI) and Modelling and Simulation (M&S) are key technologies for enabling these capabilities, showing considerable progress in recent years. AI technologies become increasingly proficient in cognitive tasks such as situation assessment, course of action planning, and teaming with humans. Advances in simulation achieve higher levels of realism in representing physical and behavioural elements of the battlefield. Furthermore, there is growing synergy between the technologies, as seen by applications of e.g. generative modelling and reinforcement learning.*

In this paper, we present an overview of the multi-faceted and interdependent roles of AI and simulation for empowering decision support capabilities in military decision-making. Based on an analysis of decision support, we assess upcoming AI technologies and their potential impact. Subsequently, we identify

considerations for military M&S platforms for exploiting these technologies. For illustrative examples, a case study is used in the domain of tactical air command & control.

The aim is to provide insights within the community on the impact of current AI advancements in enhancing military decision support, and the corresponding needs of M&S technologies, building upon existing R&D efforts in the field.

Discussion: Dr VAN OIJEN provided a general overview of AI/ML techniques. The current limitations to AI's support of C2, e.g. the lack of strategic thinking and ethical considerations, will require a gradual evolution of AI in this context. The presenter offered the following considerations for AI integration into mission simulations based on challenges the team encountered:

- AI training environment
  - Scenario generation for smart learning strategies
  - Computational demands for efficient learning, e.g. HPC
- Data availability and quality
  - Synthetic data generation
  - Data augmentation
- Digital twin environment
  - Model transfer and interoperability
- Mission context
  - Mission-specific knowledge and rules
  - Retraining and fine-tuning models

When asked about the impact of LLMs' tendency to try to "please" the questioner by providing a made-up answer when no answer exists or picking an answer they think the questioner will like best, the presenter reported that the team considered challenges of LLMs "hallucinating."

There was insufficient time during the session to ask the following questions:

1. What are the risks of using generative AI to generate appropriate mission environments and behavioural activity?

2. Does the authors' definition of digital twin include platform instance specific data?

Recommendation: See summary recommendation 3. The alignment of AI/ML techniques to components of the decision-making process could be a valuable input to the proposed tasks.

3.5 Session 9 – Human Behaviour Modelling

Session 9 Chair – Mr Wim HUISKAMP (Scientific Advisor, NMSG, TNO Defence Research)

3.5.1 Paper #26 - A Reference Model to Facilitate Collaboration in Human Behaviour Representation for Security and Defence Simulation Modelling; Dr Susan AROS, Naval Postgraduate School

Paper Abstract: *Complexities of human behaviour can cause unexpected outcomes in security and defence operations. In order to better understand the range of potential complications and outcomes it is essential to leverage constructive simulation in order to move beyond the real-world limitations of time and space. To date, however, there has been limited representation of the nuances of effects, both kinetic and non-kinetic, on human behaviour in defence modelling and simulation. This ongoing challenge is being addressed by researchers* *across multiple disciplines and countries, but collaboration is hampered by different conceptual approaches, knowledge systems, terminology and, in some cases, different definitions for the same terminology. In this paper we cover essential definitions for constructive simulation modelling of human behaviour and present a context-agnostic agent behaviour reference model developed during NATO MSG-198. In the presentation of this model we establish common concepts and terminology to aid in cross-disciplinary discussions of human behaviour modelling in constructive simulations. We then present examples to demonstrate how this reference model can be used to establish common ground for human behaviour modelling across different simulation disciplines, approaches, and application areas.*

Discussion: Dr AROS presented the following challenges in human behaviour representation (HBR) for defence M&S that make collaboration challenging:

- HBR is approached by a wide range of disciplines, which results in different approaches, use of language, knowledge systems, and perspectives.
- Increasingly complex environments require more nuanced HBR models.
- Non-kinetic effects on human behaviour are often overlooked.
- Much of the work uses bespoke, single-use models.

The goal of MSG-198 was to create a common reference model for communicating about human behaviour models as illustrated in Figure 19.

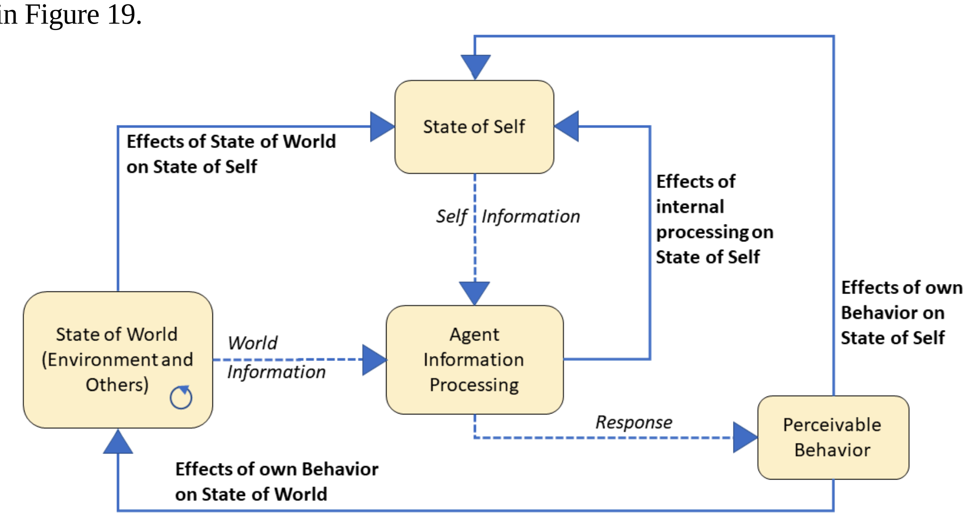

Figure 19. Agent Behaviour Reference Model

Each agent in a simulation has its own state of self. Different agents may have different techniques for agent information processing. The model was tested by rendering the Influence and Infrastructure Combat Model (IICM) and the Workbench for refining Rules of Engagement against Crowd Hostiles (WRENCH), which the presenter briefed in detail. This reference model simplifies and clarifies conversations about how individual simulations represent behaviour. This work provides the foundation for MSG-222, "Representing Human Behaviour and Decision-Making in M&S." It is also intended to drive a discussion about data including availability and format.

Perceivable behaviour is probably represented in a federated simulation in the data exchange model. However MSG-222 won't go as far as standardized external APIs and integration into federated simulations, that is not the immediate direction of the work.

There was insufficient time during the session to ask the following questions:

1. Does authenticity = validation? Abstraction = resolution?

2. The description of the IICM immediately follows the characterization of the differences between human behaviour models for training vs analysis. But IICM *"… can be used as a sandbox environment in a training and education context, and as an analysis tool."* This appears to obviate the characterization of differences.

3. Is WRENCH for training or analysis?

Recommendation: See summary recommendation 4.

3.5.2 Paper #27 – Capturing Variability in Human Capability in Mission Models for Human Autonomy Teams; Dr Marina RANTANEN-MODEER, Saab Dynamics AB

Paper Abstract: *Autonomous functions are rapidly making their way into all stages of human problem-solving. As a matter of fact, they can easily be envisioned as independent and equal team-members in a plethora of domains within the foreseeable future. The Human-Autonomy-Team (HAT) is a concept for describing collaboration between human operators and machine-based elements – a paradigm-shift in which machines no longer are mere tools, but rather helpful partners in complex problem-solving. HAT has gained attention in the defense domain where research efforts are spent on a wide spectrum of perspectives. This paper presents work-in-progress and sets out to investigate how to prepare for models that can describe the dynamics between humans and autonomous systems by introducing reliable approaches to modelling the inevitable variability in human capabilities. The ultimate goal is to enable mission autonomy, which involves agent-agnostic, independent decision-making in dynamic, operational environments that are challenging to predict. In other words: this work aims to lay the foundation for planning the unplannable through effective models.*

Discussion: Dr RANTANEN-MODEER began by stating the primary research objective of this work as defining methods for and initial results of capturing variability in human capability in mission models for human-autonomy teams. The presenter argued that variability of human capability in mission planning for HATs is of principal importance as it directly shapes mission success and adaptability in complex, dynamic operational environments. Moreover, recognizing human diversity across cognitive, emotional, and physical capabilities supports tailored task allocation, optimized human-machine collaboration, and flexibility to adapt to evolving mission conditions. The presenter pointed out that this work is conceptual at this stage.

Figure 20 gives an overview of the concept of mission autonomy across humans and systems. Dr RANTANEN-MODEER described mission autonomy as involving coordinated and goal-oriented tactical reasoning built upon the capabilities and assigned mandate of entities. Mission autonomy requires each tactical entity to have the ability to translate perception and knowledge of its mission, environment, resources, and bounds into an adaptive course of action aimed at a common goal. The presenter felt that autonomous functions are inevitable in the battlefield since adaptability to dynamic environments can yield accuracy and safety, giving weight to the need to model the dynamics between autonomous functions and humans. An aspect of this capability is to become able to plan for unexpected or unpredicted situations. Such systems have a capability to reason and act according to doctrine, much as humans are trained to doctrine but also require an ability to adapt and exercise initiative in the face of different conditions.

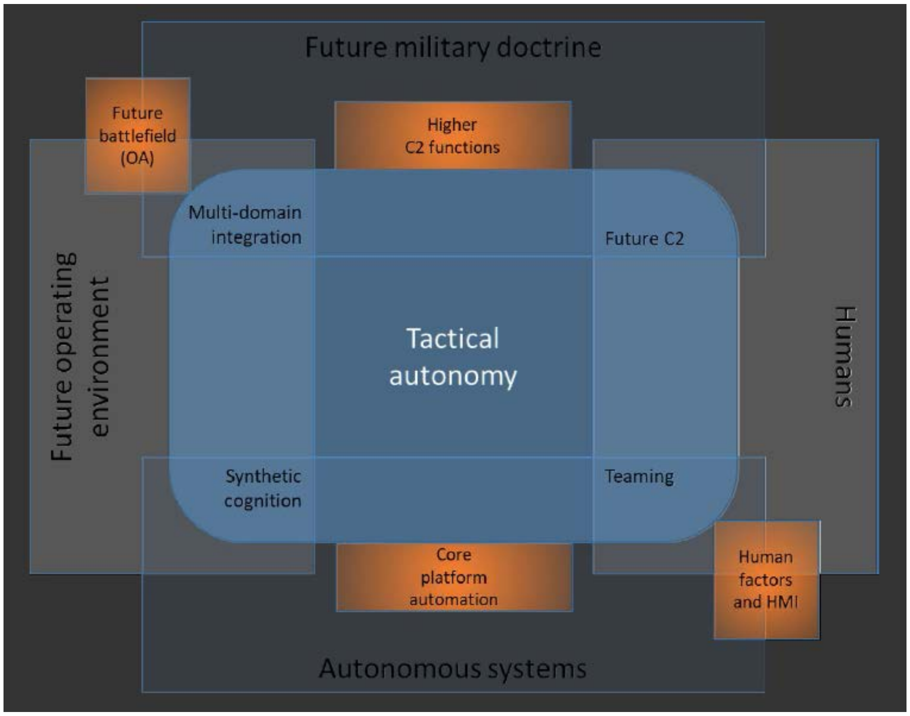

Figure 20. Mission Autonomy: Humans and Systems in Future Operating Environment and Future Doctrine.

Dr RANTANEN-MODEER emphasized that humans must retain absolute authority over the systems. There is a need to determine where systems can aid humans to improve mission success. It is interesting, though, to recognize that machines are easier to model than humans. There are various ways to represent human capabilities, such as agent-based modelling (e.g., refer to paper #26), systems dynamics modelling, Monte Carlo simulations, and Bayesian decision networks. Team dynamics add complexity since there are several different relationships that have to be considered; i.e., humans-to-humans, humans-to-systems, and systems-to-systems. In their approach, humans are modelled as a special class of sensors.

The presenter described interpretation of mission intent as a key consideration, where that interpretation needs to generate appropriate tasks, action, and expected effect(s) from the communicated (explicit) intent and unvocalized (implicit) intent (based on personal expectations, military expectations, and cultural expectations). Team task execution involves a task planner and an execution supervisor, with task prioritization and task coordination. The presenter described a concept of synthetic cognition with interplay between aspects of observation and orientation and aspects of decision and action, operating at higher and lower cognitive levels.

Dr RANTANEN-MODEER described an experiment framework involving a legacy manned simulation and Command Modern Operation. Human operators are monitored to record cognitive load, stress levels, and environmental factors. Data are passed to a centralized AI system to represent the humans as sensors. Based on information from the human capability variability model, the AI system provides real-time feedback to each operator, offering suggestions to manage stress and optimize performance. The approach was described in the presentation and in the paper, identifying both current observations and limitations. Future work involves investigating the use of large language models for interpretation of intent and mapping human response to a wider range of inputs.

In discussion, the audience brought up the consideration that machines have different failure modes (such as software errors) than humans do. These need to be characterized and included in system designs. The audience indicated interest in obtaining data from this work for possible inclusion in other simulations. The audience also pointed out that the presentation implied human decision making and autonomous system decision making occur on the same time scale when, many felt, the scales could be (and are generally expected to be) significantly different.

The reviewers believe this is important work, having argued for the past eight years that the rapid operational transition to human-unmanned system teams has advanced far faster than the ability to model such teams effectively. There are clear differences in human system performance and robotic/autonomous system performance, but the M&S community has not received the resources needed to be able to model those differences effectively. Instead, representation of humans continues to be little different from representation of robotic systems. Serious attention (as in, funded projects) is needed in this area.

Recommendation: See summary recommendation 1 and 4.

4.0 SUMMARY AND RECOMMENDATIONS

The following is a summary of the presentation by the technical evaluators at the end of MSG-207. The items contained here are solely the opinions of the technical evaluators and are not intended to reflect any form of official position.

The keynote and invited presentations provided multiple perspectives on the NMSG scope.

- Mr MCALINDEN not only covered the scope of the technologies that intersect with our scope, but advocated for the integration of creative professionals (e.g., entertainment, story-telling) into the process of ideating future situations and technologies (strategic advantage of imagination). There is a major pivot in determining how to quickly train and educate the warfighters.

- Mr TEMPALSKI highlighted CMSO's initiatives to support NATO, including
  - Creating a centralized model repository of releasable models
  - Creating more opportunities for foreign national students at the USAF Academy
  - Updating the Warrior Prep Center (WPC)
  - Opening up 2 seats for each NATO nation at the annual USAF M&S Summit to be held the first week of May in San Antonio, TX.

- VADM (ret). RONDEAU issued a call to action to learn from recent world situation surprises to project our thinking into the future, i.e., to be better before we need things, while reminding us that we have a duty to prepare our future warfighters for future challenges and surprises.

- Mr KALLFASS demonstrated the integration of multiple AI/ML technologies, which is directly on theme for the Symposium. The Way Ahead section of his paper could provide valuable input to the recommendation 3 tasking.
  - Continued Training and Evaluation
  - Scalability and Real-World Testing
  - Human-in-the-Loop Integration
  - Versatility in AI Applications
  - Ethical and Legal Considerations

- Testing and Validation
- Adaptation to Evolving Technologies

Recommendations:

1. Because the topic of the symposium changes every year and presentations are often on work in progress, attendees may not receive the full benefit of the final outcomes of the projects presented. Additionally, the symposium's focus on new and emerging technologies guarantees the topic field is moving very fast, making continuous engagement by attendees highly beneficial. Establish a virtual mechanism to maintain momentum on the topic including follow up detailed questions to presenters where all attendees can view answers, such as a wiki or knowledge base. It could also be a mechanism for sharing results of ongoing work. This is recommendation #3 from MSG-171, the 2019 NMSG Symposium.

2. Collect and consolidate NATO uses of M&S standards.

3. Consider establishment of an exploratory team or specialist team to develop a white paper to explore the intersection of:
   a. M&S domains: testing, training, analysis, experimentation, etc.
   b. Use cases for each domain that could benefit from the application of AI / ML
   c. Specific AI/ML techniques most applicable to the use cases
   d. Risks and challenges associated with applying specific types of AI/ML to use cases
   e. The intersection of MBSE system models with military battlespace simulations

   We understand this area has been examined previously from various perspectives. The first task of this team would be to discover and consolidate progress to date.

4. Consider as a future theme rearchitecting simulations to integrate emerging technologies:
   a. AI / ML
   b. Cyber / EW
   c. HBR