Human performance interoperability: Using cognitive and behavioral analytics in immersive, adaptive working environments

Whit Missildine[1], Christopher Clift[2]
[1]Chief Learning Officer, Viion, Ltd., Oakland, CA, USA
[2]Director of Training Capability, Babcock International, Helensburgh, UK

Abstract —While technological capability and strategic interoperability are often emphasized as the critical elements to address emerging threats, human performance interoperability is too often overlooked. Training and human performance assessment will need to evolve to meet the challenges of the future military operating theatre, by getting more meaningful data, more rigorous analytical frameworks, and more effective and efficient training environments. Babcock International Group PLC and Viion Inc. collaborated to combine their individual areas of research and create a synthetic, multiplayer virtual submarine environment that challenged users to manage a complex collision event. To assess performance, we developed an analytics engine that synthesized multiple human performance data inputs in line with a universally applicable model of human learning and development, expertise and experiential learning. We will present initial findings from user trials with Royal Navy personnel and lessons learned regarding the use of interoperable human performance analytics in the adaptive training context.

1 Introduction

Gaining agreement on a "broad definition of interoperability" has proved difficult, because the definition can shift depending on the situation, degree and level of operation (strategic, operational, tactical, technological). However, a RAND report on interoperability sponsored by the US Air Force, put forward the following definition: "The ability of systems, units, or forces to provide services to and accept services from other systems, units, or forces, and to use the services so exchanged to enable them to operate effectively together." [1] Within this broad definition, interoperable complexity emerges at multiple levels, and each level may have bespoke solutions, making, for example, inter-platform, inter-task force or group, inter-service, and international data non-transferable.

Furthermore, interoperability commonly focuses on the shared equipment, technologies, and training infrastructure that help coordinate joint action to address complex threats and adapt to a changing international defence landscape. For example, effectively sharing data is a critical factor in achieving inter-service interoperability. When systems-level interoperability is addressed, the focus is often on the inter-service transferability of data regarding equipment and assets. A military platform may be generating data about its speed, environmental sensors, or surface damage, and that data needs to be compatible with other systems in a joint operation. But an even more critical factor is often overlooked: what about the people operating the platform? While the interoperability debate usually emphasizes strategic and technological capability as the critical elements to address emerging threats, generating interoperable human performance data is too often ignored.

Several factors may account for this deficit. First, evaluation of human performance is often left to subjective assessment by specific training personnel and protocol, who all have their own methods of assuring proficiency.

Second, human behavior is often considered inherently unpredictable. Emphasis is placed on buffering against decision making deficits or human error by creating tools and technologies that remove human factor risk rather than implementing enhanced training and readiness at the level of human performance.

This leads to the third factor: funding. Demonstrating added value for military investments is easier with technological solutions that afford hard data on improvement. But funding is often disproportionately allocated for, for example, a sensor that increases the accuracy of threat detection, rather than training the user on how the sensor's data should be interpreted, communicated and deployed to address the detected threat. While basic competencies are part of standard protocol in these domains, accounting for skill and experience levels, complex and atypical threats situations, and handling equipment and technology failure require a deeper understanding of human cognition and behavior. Funding and resources for this kind of deeper dive is often lacking. In this example, enhancing sensors provides a clear added value in threat management that creates an obvious route to funding. But how the data is interpreted, communicated and deployed (i.e. enhancing human performance), while equally valuable, often gets a disproportionately small share of funding resources.

These three factors combine to make human performance a highly underserved aspect of the interoperability debate, and yet, from our perspective, it is the key element that sits at the center of it and potentially links all of the other elements together.

1.1 Critical areas of need

The future military operating theatre will see a number of unique, large-scale shifts that will place new demands on operational readiness. The Future Operating Environment 2035 [2] document lists a number of emerging threats that will radically alter how warfare in conducted. Rapid urbanization, electronic warfare, climate change, humanitarian crises, and the rise of non-state actors all create conditions for new forms of warfare, which drive new training needs. These anticipated shifts will require a radically different approach to training and should prompt us to rethink what human performance interoperability means in a connected age. Talk of automation and reduced risk to ground troops has called into question the future need for on the ground readiness. Autonomous warfare that relies on unmanned aerial and ground cargo vehicles to support troops in combat zones will change the nature of resupply efforts. But they will likely never replace the needs for troops on the ground. Furthermore, the introduction of sophisticated technologies will not obviate the need for critical human decision skills. Instead, new technologies, operational environments and the situational stresses they pose may actually increase the need for high cognitive agility and adaptability to protect against what UK General Mark Carelton-Smith called a “darkening geo-political picture.” The question is: are our training and learning models ‘fit for future’? [3]

Human performance interoperability will require adaptive and accelerated training programs focused on human agility in decision making, as well as new methods for collecting, assessing, synthesizing and sharing human performance data. However, achieving these goals raises a number of important questions: how do we generate not just bigger data, but more meaningful data on human performance? What analytical frameworks are best suited to adapt to a future military operating theater? How can we enhance our current synthetic training environments to reduce interoperability risk? How do we create interoperable proficiency standards across varying strategic, operational, and tactical contexts?

More specifically, we see three areas of need that are essential for achieving effective human performance interoperability:

1. Standardized training procedures and environments that can be delivered to the point of need.
2. Shared assessment and evaluation protocol that create a baseline for joint proficiency assurance.
3. Adaptive methods that can reveal vulnerabilities in joint operations and adjust to player behavior.

1.1 Human performance evaluation using synthetic environments

One of the most important advances in training in the last few decades is the emergence of immersive virtual technologies, which allow for complex, remotely delivered, experiential, multiplayer training scenarios which can be created an increasingly affordable price point. The UK Army’s Future Collective Training System, for example, highlights the importance of simulation and virtual reality technologies to drive future training programs. However, they emphasize that “finding the right mix of live, virtual and constructive training is key, with the boundaries between the three becoming increasingly porous, effectively forming a single domain.” [4] The Royal Navy’s Maritime Training Strategy mirrors this view in its desire to exploit immersive, technology-enabled collective training to deliver “a maritime training system that prepares people, units, platforms and formations to deliver successfully the full spectrum of maritime operations.”

Moving training into synthetic environments and creating an “adaptive blend” of live, virtual and constructed training tools has a number of advantages over traditional training, including a reduction in instructional resources, more realistic and flexible scenario design, delivering to the point of need, and, most importantly for us, a way to create standardized assessment techniques with get rich feedback and analytics.

Synthetic environments are especially suited for training situations which are impractical, difficult, dangerous or expensive to reproduce in an operational environment. There are many potentially dangerous situations that trainees may only encounter infrequently, if at all, but when they are they need to be dealt with efficiently to avoid serious consequences. These environments can be used to present trainees with such unusual scenarios in a repeatable and controllable fashion. These scenarios are ones in which human performance and decision making are most critical, and ones that do not often have a technological or systems-driven solution.

Simulation in synthetic environments provides a consistency of training and assessment that is not possible in the operational environment, enabling standardized objective assessment and measured against a common standard. These environments also allow us to track and record human performance data in unprecedented ways. Better use of these environments as part of an enhanced training capability can significantly increase learning effectiveness and reduce skill-fade. Full optimisation enables individual and collective practice and repetition to further reduce skill-fade. In the context of interoperability, the use of synthetic training environments will become increasingly important through their ability to standardize processes, practice joint missions, and assure mutual proficiency. And yet, assessment and evaluation models have yet to catch up with these opportunities.

With more visibility into behavior and decision making, including better tracking systems, controlled scenarios, and real time feedback mechanisms, synthetic environments often end up generating large data sets. But as generating large data sets becomes easier, a new issue arises: how do we generate not just more data, but *more meaningful* data. This issue becomes even more important when the methods scale up in the context of interoperability. So the full question becomes: How do we generate more meaningful data on human performance that can be effective in interoperable contexts?

2 The Adaptive Learning Approach

To create the most effective scenarios, a learning approach is needed that utilizes the full affordances of multiplayer synthetic environments to maximize learning outcomes. "Adaptive Training" is a useful umbrella term for a set of principles that go beyond classical "skill and drill," practice-based learning methods. This adaptive approach focuses on challenging participants to think holistically, to innovate solutions to novel problems and to make decisions that can go beyond strict reliance on procedures and protocol. Instead of narrowing in on rote memorization and recall, procedural adherence, or task compliance, adaptive training focuses on 1) how we help learners move through developmental stages of expertise, 2) how we customize training to different learning styles, and 3) how to leverage analytics and feedback to create challenging scenarios that adapt to learner progress.

Critical to the future success of human performance interoperability, particularly in synthetic environments, will be the need for common metrics and evaluation criteria rooted in a universally applicable model of human learning and development. Getting the learning model right enables an adaptive approach to data capture and assessment that is both platform agnostic to the technology, role agnostic to the organization, and domain agnostic to the targeted learning outcomes.

Simultaneous with the advent of virtual reality has been an explosion of research in cognitive and neuroscience that have begun to uncover the common mechanisms behind how we make decisions and how we develop not just experience but expertise. Our work combines several lines of research that address various aspects of learning and expertise development in complex situations.

First, we build on the Dreyfus 5-stage model of expertise, along with naturalistic decision making and macrocognitive models of expertise to better understand the continuum of skills needed to remain effective in high stakes, high uncertainty operating environments [5,6]. Next, we use Piaget's model of disequilibration [7] and the Predictive Processing Model [8] as a framework for driving change through the developmental trajectory. And finally, we employ Kolb's Experiential Learning Model [9] to address individual learning styles, and create an adaptive model wherein scenarios can adjust to player behavior and performance. Leveraging these learning models, we built an intelligent data evaluation and assessment framework that allows us to integrate and make sense of large data inputs and outputs.

2.1 Submarine Case Study

In February 2018, Babcock International Group and Viion, Inc. collaborated to create a synthetic, multiplayer, immersive virtual submarine environment for the Royal Navy that would accelerate expertise, reduce at-sea training burden, and improve platform availability. The adaptive virtual training environment was meant to help learners develop well-practiced familiarity with formal procedures as well as rare or unforeseen events to better manage unpredictable sources of risk. Our goal was to create scenarios that would push decision makers beyond proficiency assurance by using atypical events to train for higher level sensemaking and leadership capabilities in uncertain conditions.

As part of the scenario design, we had to determine the appropriate metrics that would capture baseline capability, procedural knowledge, tactical competency and expert-level decision skills. As such, we needed an assessment model that would act as both a measure of basic competencies while also driving behavior change and stretching the limits of expertise.

The scenario we designed challenged a submarine Ship Control Officer of the Watch (SCOOW) with a complex collision event. The scenario had multiple intervening factors. These factors included hazardous distractions that tested knowledge of basic EOPs; injury to crew members which created heightened urgency and additional noise in the system; and conflicting sources of information that challenged the participant's sensemaking abilities. To measure performance, we developed an analytics engine that integrates inputs from 1) pre-operational diagnostics, using basic experience indicators and Kolb's experiential learning styles model, 2) real-time facilitation ratings based on Royal Navy proficiency assessment standards, and 3) post hoc assessments based on Dreyfus's 5-stage model of expertise.

We synthesized these metrics into a single user profile that can evolve over time, across learning modules, and along the Dreyfus developmental trajectory as the learner's career advances and new modules become available. Furthermore, while the system can adapt to an organization's unique tasks, protocols and procedures, the underlying framework relies on universal models of learning and adaptive approaches that allow for shared, interoperable training methods. While this scenario is specific and unique to a submarine SCOOW, the assessment paradigm is scalable and adaptive to varying operational contexts.

2.2 Results

The initial data has allowed us to compile and analyse the cognitive and behavioral outcomes demonstrating to the Royal Navy individual and collective assessment can be data driven. Further validation and assessment metrics are being developed to give the correlation between subjective assessments with objective data outputs.

A non-technical skill (NOTECH) list typically comprises: co-operation, leadership and management, situational awareness, and decision making. Many of the positive and negative assessment criteria associated with these have been built into the analytic engine, but broken down further to reflect more granular cognitive dimensions than simply "decision making." Some of the more granular cognitive assessment metrics include:

- Sensemaking
- Coping with Complexity
- Sorting and Prioritisation
- Emotional Regulation

The correlation between the cognitive and behavioral elements and the individual's level of expertise and learning style at any given time during the scenario comprise the full picture of an individual's performance, and this is tracked as the exercise unfolds. Other performance data, such as communication, is generated dynamically and displayed in a graph format against scenario elapsed time.

The results below are derived from the initial test datasets and are included to show expertise level transition and trajectory during a sample five-minute time frame. Figure 1 shows a novice user, figure 2 an experienced user, and figure 3 an expert user. The complex collision event occurs at the 120 second marker.

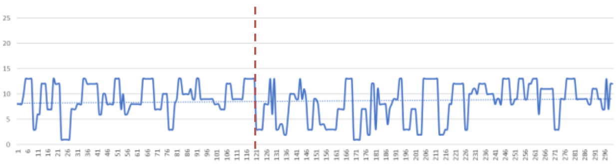

Fig. 1. Cumulative Novice Performance

Fig. 2. Cumulative Experienced Performance

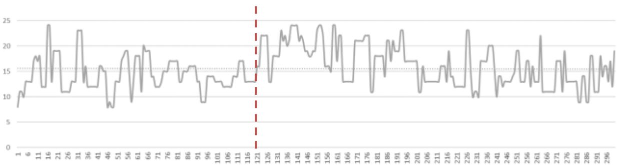

Fig. 1. Cumulative Expert Performance

Figures 1, 2, and 3 shows skill progression on the X-axis on a scale of 0 to 25. The progression is through the five levels, each of which has five subdivisions. The Y-axis shows time, with data points sampled at one second intervals. These results indicate a weakening of performance for novices when confronted with high uncertainty events. The results show high volatility for experienced users when handling the high uncertainty event. The expert user demonstrates little change in performance during the high uncertainty event.

The second set of results below are also derived from the initial test datasets and are included to show learning style transition and trajectory over the same five-minute time frame. Figure 4 shows a novice user whose natural learning style was predominantly Reflector. Figure 5 shows an experienced user whose learning style was predominantly Theorist, and figure 6 shows an expert user whose predominant learning style was Activist. The complex collision event occurs at the 120 second marker.

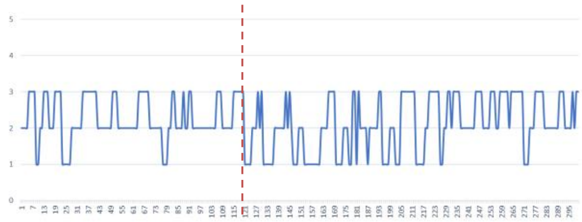

Fig. 4. Learning Style Novice Performance

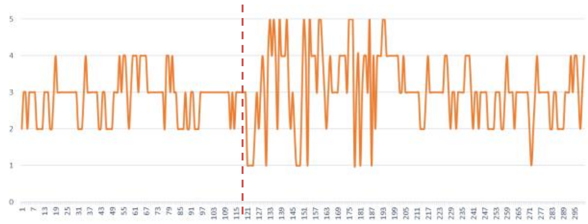

Fig. 5. Learning Style Experienced Performance

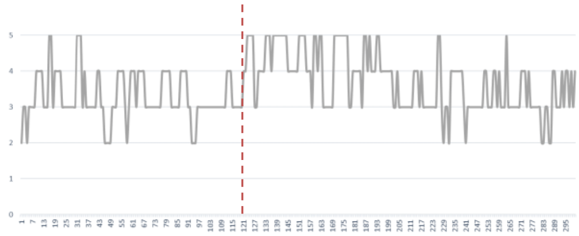

Fig. 6. Learning Style Expert Performance

Figures 4, 5, and 6 shows how each individual's learning style adapted during the scenario. The X-axis shows Reflector – 1, Theorist – 2, Balance – 3, Pragmatist – 4, and Activist – 5. The Y-axis shows time, with data points sampled at one second intervals.

More results are being assessed as the dataset grows but the initial takeaways are in a role like the SCOOW much of the activities are routine and procedural so the averages in Figures 1, 2, and 3 are close, respectively 8.7, 12.8, 15.5. this indicates during normal or routine operations we are mainly assessing competence not experience or expertise. However, during the two-minute period immediately after the complex non-routine event the same averages were: 7.4, 13.9, 17.6 which gives us a clearer assessment of both experience and expertise.

Initial indications are that changes in learning style during the scenario are different for the three types individual. During normal or routine operations, we see all three centering around Balance with a significant change at the complex collision event. The split for Figure 4 is Reflector – 19%, Theorist – 46%, Balance 35%, Pragmatist – 0%, Activist – 0%. The split for Figure 5 is Reflector – 4%, Theorist – 20%, Balance 51%, Pragmatist – 20%, Activist – 5%. The split for Figure 6 is Reflector – 0%, Theorist – 6%, Balance 48%, Pragmatist – 32%, Activist – 15%.

3 Initial Conclusions

Creating an end to end learning model capable of generating the full spectrum of human performance data in the interoperable contexts is key to scalability, resilience, and sustainability of defence outputs. Leveraging these learning models and making them central to our synthetic and live training environments will deliver initial knowledge, refresher training and full assessment for individual and collective training.

Assessing human performance in the interoperable context requires two key ingredients: 1) the assessment model must conform to a universally applicable model of human development and expertise, and 2) it must have as its core benefit the skill of adaptability. Addressing these two key ingredients reveals a gap in the current protocol, which lacks sufficient detail and meaningful data on the cognitive dimensions of human performance.

By focusing on the cognitive dimensions of human performance assessment, the model provided a framework for evaluating previously overlooked dimensions of human performance. The model does this by focusing on three factors: a developmental, stage-based model of expertise; a motivational model of adaptive reasoning; and a learning styles model that accounts for individual differences. By capturing all stages of the learner's developmental trajectory we can anticipate what the needs will be at the next stage in the training process. By incorporating a learning model that motivates adapting reasoning, we can prepare learners for a wider range of situational demands. By adjusting to the learner's unique style, we can accommodate a larger set of trainees and target specific developmental needs. Perhaps more importantly, the model demonstrated the possibility for scalability and how it could be applied to the full spectrum of military platforms.

While traditional training and assessment methods continue to be valuable for assessing performance on routine tasks, procedures and protocol, they are not well suited for a deep understanding of performance during high uncertainty events, where cognitive skills associated with sensemaking and decision making are critical. In our initial trials, we were able to differentiate performance among varying levels of novice to expert operators using a more granular approach to assessment along key cognitive dimensions of expertise.

In the human performance interoperability context, creating adaptive models for assessment rooted in universally applicable models of human development is the key to ensuring readiness is not limited to a single military service or to one country's defence output. The future demands on human performance interoperability will require a greater understanding of how operators make decisions in high uncertainty contexts, using new training tools and paradigms to ensure proficiency and operational readiness.

References

[1] M. Hura, et. al. Interoperability: A continuing challenge in coalition air operations. RAND Monograph Report (2000)

[2] Ministry of Defence. Future Operating Environment 2035 (2014)

[3] Ministry of Defence. British Army set to redefine warfare with joint Autonomous Warrior. https://www.gov.uk/government/news/british-army-set-to-redefine-warfare-with-joint-autonomous-warrior

[4] UK Army. Collective training transformation programme. Market Engagement Paper 1, p.6 (2018)

[5] H. Dreyfus. Intuitive, deliberative and calculative models of expert performance. In C. Zsambok & G. Klein (Eds.) *Naturalistic decision making.* (Lawrence Erlbaum Associates, 1997)

[6] G. Klein. A recognition-primed decision (RPD) model of rapid decision making. In G. Klein, J. Orasanu, R. Calderwood, & C. Zsambok, (Eds.), *Decision making in action: Models and methods* (Ablex, 1995)

[7] J. Piaget, J. (trans. A. Rosin). *The Development of Thought: Equilibration of Cognitive Structures*. (Viking, 1977)

[8] A. Clark. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Beh. Brain Sci.*, 36, (2013).

[9] A. Kolb & D. Kolb, D. Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning & Education,* 4 (2005).

Author/Speaker Biographies

Dr. Whit Missildine has a PhD in Social Psychology and is the Chief Learning Officer at Viion, Inc. He has led research, design and implementation efforts on numerous immersive learning engagements and had delivered peer-reviewed publications and presentations on simulation-based training, immersive learning, accelerated expertise, and learning experience design.

Christopher Clift is the Director of Training Capability with Babcock. For over 35 years he has worked in commercial aerospace and defence developing immersive learning environments. He is pursuing a PhD through the University of Strathclyde and working with the Royal Navy, British Army, Government, academia and industry to evolve adaptive training in defence.