Are Current Modeling Architectures Viable for Rapid Human Behavior Modeling?

Margery J. Doyle; Antoinette M. Portrey
L-3 Communications, Link Simulation and Training
Wright-Patterson Air Force Base, Ohio
Margery.Doyle@mesa.afmc.af.mil; Antoinette.Portrey@mesa.afmc.af.mil

KEYWORDS
Cognitive Modeling, Modeling Architectures, Constructive Adversaries, Adaptive Agents

ABSTRACT

The Department of Defense is developing plans to transform training and education; providing more dynamic, capabilities-based training programs in support of national security requirements associated with today's strategic environment. Capabilities-based training must facilitate the warfighter's ability to respond quickly and adapt effectively when faced with rapidly evolving, asymmetric warfare. Training tools such as future threat generation systems, modeling techniques, and performance measurement methods themselves must be flexible enough to cultivate responsiveness and adaptability in current and future warfighters.

The development and execution of adaptive threat generation systems and rapid modeling techniques within applied research and training environments poses many methodological and integration challenges. The Air Force Research Laboratory, Warfighter Readiness Research Division (711 HPW/RHA) is assessing the critical issues facing present day threat generation systems, threat models, and the extent to which current modeling frameworks/architectures can provide military training with accurate/credible models of human behavior. The objective is to explore the efficacy of incorporating real-time performance data into architecture(s) to develop models (e.g., realistic constructive adversaries) for use in a threat generation/adaptive training systems. Also, the possible types of assessment criteria (e.g., verification and validation) needed to define levels of successful integration.

This paper will briefly examine the critical issues facing future warfighters, present a high-level summation of the current research and practices associated with threat generation/adaptive systems and their current use of modeling techniques, and will discuss salient lessons learned. Finally, the team will suggest directions on how these tools can be used to better prepare the warfighter to answer the dynamic challenges of future warfare. This research can serve as a foundation in the development of more adaptive, capabilities-based training tools (e.g., less-scripted, more dynamic training scenarios with realistic constructive forces) for use in training and ops communities such as Live Virtual Constructive (LVC) environments.

ABOUT THE AUTHORS

Margery Doyle is a Cognitive Research Scientist with L-3 Communications Link Simulation and Training at the 711 HPW/RHA in Wright-Patterson Air Force Base, OH. Currently she is developing cognitive engineering and research methods to support development, evaluation, integration, and utility in the use of cognitive models, agents, and architectures in adaptive live virtual constructive training environments. She earned her M.A. in Experimental Psychology with a certificate in Cognitive Psychology from the University of West Florida in 2007. She has also completed work towards a PhD in Cognitive Science. Margery is currently co-editing a special issue of Cognitive Systems Research on Stigmergy to further the integration and use of stigmergic processes and techniques in areas such as pervasive computing, multi-aircraft control, and social network analysis.

Antoinette M. Portrey is a Senior Scientist with L-3 Communications Link Simulation and Training at the 711 HPW/RHA in Wright-Patterson Air Force Base, OH. She leads the Training Research Team in the development of training research methodologies and technologies, including the Performance Evaluation Tracking System (PETS). Ms. Portrey is also co-editor for "Individual and Team Skill Decay: The Science and Implications for Practice" (In press). She completed her M.S. in Applied Psychology - Human Factors from Arizona State University in 2005 and an MBA in Project Management from Capella University in 2008.

Are Current Modeling Architectures Viable for Rapid Human Behavior Modeling?

Margery J. Doyle; Antoinette M. Portrey
L-3 Communications, Link Simulation and Training
Wright-Patterson Air Force Base, OH
Margery.Doyle@mesa.afmc.af.mil; Antoinette.Portrey@mesa.afmc.af.mil

INTRODUCTION

The significant transformation of today's strategic environment has had an extensive impact on the warfighter and how the Armed Forces prepare for combat operations. In the 20th Century, the Armed Forces were a requirements-based military preparing the warfighter using traditional methods that have been around for decades. However, the 21st Century mission mandates a more modular, interoperable, extensible, scalable, flexible, adaptive mindset (United States Air Force, 2004). Therefore, the Department of Defense (DoD) is planning to transform training and education; providing readily adaptable, more modifiable, dynamic, capabilities-based training programs in support of national security requirements associated with today's strategic environment. These capabilities-based training programs must facilitate the warfighter's ability to respond quickly and adapt effectively when faced with rapidly evolving, asymmetric warfare. In turn, training tools such as future threat generation systems, modeling techniques and performance measurement methods must be flexible enough to cultivate responsiveness and adaptability in current and future warfighters.

The design, development, and execution of adaptive threat generation systems and the use of rapid modeling techniques within applied research and training environments pose many methodological and integration challenges (Office of Aerospace Studies Air Force Material Command, 2005; Wang, Tolk, Wang, 2009). The Air Force Research Laboratory, Warfighter Readiness Research Division (711 HPW/RHA) is assessing the critical issues facing present day threat generation systems, models, and the extent to which current architectures can provide military training with accurate/credible models of human behavior. One objective is to explore the efficacy of incorporating real-time performance data into an architecture to develop entity models (e.g., realistic constructive adversaries) for use in an adaptive threat generation system. By default of this process another objective is to research the possible types of assessment criteria (i.e., verification and validation) needed to define levels of successful model integration within a threat generation system.

Relevance

The Strategic Plan for the Next Generation of Training for the Department of Defense (Office of the Under Secretary, Personnel & Readiness, 2010) and the Developing an Adaptability Training Strategy and Policy for the Department of Defense (Freeman & Burns, 2010) recommend the establishment of a robust well focused research and development program aimed at further defining: training in support of adaptability, and the measurement of operator adaptability and team performance.

The Strategic Plan (USD [P&R], 2010) states the need for adaptive training systems to be able to:

- Achieve sufficient level of technical and operational realism (Realism/Validity).
- Use common applications, references, and operational capabilities (Commonality).
- Be rapidly scalable (small team to joint task force [JTF]) (Scalability)
- Composable by users without the need of specialized or proprietary skills. (Modularity/Usability)

The strategic plan mentions that these rapidly adaptive training systems need to be able to synchronize training operations in near-real time to enable realistic stimulation of sensors, replication of visual cues, and platform interactions between live, virtual and constructive (LVC) participants. Forces should also be trained in a culture of adaptability and flexibility, giving them the ability to rapidly reconfigure to address new threat profiles, while establishing the right balance between core proficiencies and just in time, adaptive training to satisfy specific mission tasking. Adaptive training administered in this format will develop a quick reaction capability through supporting individual and team augmentation of training and performance; by rapidly incorporating changes to doctrine, Tactics, Techniques and Procedures (TTPs) derived from lessons learned; and enabling our forces to take an Adaptive Stance (Grisogono, 2010). According to Grisogono, an intellectual and organizational adaptive stance can be distilled from some of these hallmarks of adaptability;

- Intelligent context-appropriate behavior
- Resilience
- Robustness to perturbations
- Flexible responses
- Agility
- Innovation
- Learning from experience
- Continuously seeking ways to test and revise assumptions
- Objectively assess what can be learned in order to improve future predictions

CURRENT STATE OF RESEARCH

Research has shown a link between experience in adaptive situations and an increase in adaptive performance (Mueller-Hanson, Nelson, & Swartout, 2009). Being exposed to the same training and gaining the same experience repeatedly (e.g., training the same task to the same standard under the same or similar conditions) may create an inflexible mindset; impeding performance in a novel situations. However, experiencing a variety of novel situations requiring adjustments to a trainee's mindset or actions appears to aid in acquiring a strong adaptive stance. Therefore, if trainees are to learn to be agile and adaptable, training events must provide the trainees with a problem for which they have not planned; a problem tailored to meet specific training needs and reinforce areas of training in need of improvement.

Institutes such as the National Research Council (NRC) (Zacharias, Macmillan, & Van Hemel, 2008) have asked the DoD to review relevant research programs that utilize and conduct research with Individual, Organizational, and Societal (IOS) modeling methods. Their goal is to evaluate the capabilities and limitations of the programs, their methodologies, and determine which methods show the most promise for use in the military domain. The NRC executive committee identified five major categories in need of being addressed:

- Modeling strategy-matching
- Verification, Validation and Accreditation
- Modeling Tactics
- Differences in modeling physical systems and human behavior
- Interoperability of models when attempting to create a federation of models

They mention that an evaluation of the model against the real world can be problematic, because real world complexities are often not properly taken into account, Subject matter experts (SMEs) are not consulted, and/or task analyses are not done thoroughly. Additionally, an understanding of the system or process a model is meant to represent is not always sufficiently defined. These shortcomings often lead to unrealistic expectations as to how much of the real world complexities should or can actually be modeled. To make matters worse, frameworks and architectures often do not scale up easily or handle changes in human or organizational behavior rapidly enough; rendering them brittle and limited in their usefulness.

Threat Generation Systems and Threat Models

Behavioral Modeling

Many of the current threat generation systems rely primarily on a single behavior modeling approach (i.e., production modeling, mathematical modeling, or Artificial Intelligence (AI)). Simple behavior is often represented by procedural scripts and while scripts may be straightforward to construct, the resulting behavior is predictable, rigid, and more often than not unable to adapt to new situations, in-turn limiting realism of training and degrading experimental efficacy. In most virtual research and training testbed environments today, adversarial agents with scripted responses in an air-to-air/air-to-ground situation are expected to laser, lock on, aim, and fire upon a pre-define adversary (usually human operators in training) when the agent first perceives any adversary in the virtual environment.

However more complex adaptive behavior modeling representations can be defined via C++/Java code, scripts (Spronck, Ponsen, Sprinkhuizen-Kuyper, & Postma, 2006), or finite state machines (Fu & Houlette, 2004); or, on the fly, by a planning system from hybrid architectures that plans within or over hierarchal task networks (Orkin, 2004; Hoang, Lee-Urban, & Muñoz-Avila, 2005); or with production rule based systems such as adaptive control of thought–rational (ACT-R) (Best, Lebiere, & Scarpinatto, 2002) or Soar (Wray, Laird, Nuxoll, Stokes, & Kerfoot, 2005). Finite state machines (including hierarchical and probabilistic), behavior trees, and simple rules (often used in gaming environments) are slightly more flexible, but can be difficult to manage as they grow in size. It is also recognized that production systems (e.g., Soar: Newell, 1990; ACT-R: Anderson & Lebiere, 1998), while providing cognitive plausibility and validity, continue to require a high level of expertise to develop.

Computational Cognitive Models

Computational cognitive models are integrated models of how humans perform complex cognitive tasks. For example: sensation, perception, knowledge, and motor action models embody the underlying theory or framework for human information processing or experience through these modalities. These models allow behavior and cognition to be simulated across a broad range of situations. Such models can predict how well systems support cognitive tasks by assessing factors such as ease of use, workload, learnability, and the propensity for a system and its interactions with a user to generate latent errors.

Unlike AI, which is designed to complete tasks faster and with fewer errors than humans, computational cognitive modeling is developed and designed to complete tasks in the same way humans are thought to complete the task; inclusive of the time constraints humans' face when completing a task. In addition to advancing the theory of cognitive processes, these models provide information about what users can/cannot and should/should not functionally do when interacting with the interface. Basically, informing the design so the use of these systems requires less initial learning on the part of users making them more tractable and usable.

Lamoureux, Bandal, Martin, and Li (2006) reviewed and categorized the functional applications of 26 modeling applications (e.g., EPIC, ACT-R/PM, Soar, IMPRINT, IPME) against 15 criteria (e.g., workload, individual/team task modeling, scenario flexibility, real-time CGF, stability). They also evaluated five multi-agent teamwork simulation tools; models designed to represent aspects of teamwork not found in most models (e.g., "off-task" behaviors interrupt and resume collaboration, multitasking, informal interaction, and geography). Since these models have the capacity to represent complex command and control interactions this class of models should be applicable to scenario, tactical planning, and execution for use in training research experimentation.

Model Scalability/Usability

Tollefson, Schamburg, and Yamauchi (2006) argue that today's Computer Generated Forces (CGF) can represent a full range of systems, control processes, and operations. CGF can range from individual combatants, to platforms representations, to multivariate constructive adversaries. All of which can interact with an individual operator and teams in a range of command modes and levels of autonomy; from manual to fully automated. CGF can even engage as adaptive generative tactical friendly or opposing forces. To add to the complexity, most CGF based models and simulations are composed of a set of products of interacting components and tools (Randolph & Sagan 2003).

While CGFs can fill many complex roles in training, CGFs require rapid development and/or re-development to quickly and accurately specify human-like behaviors in simulations. Such as the development of adversary or instructor-like responses that lead to repeatable (i.e., rote learning) or novel adaptive stance learning situations; situations introduced dynamically to elicit further skill development on a particular procedure or tactic. Or to emphasize blue force response tactics that need to be learned to a state of automaticity to build a more agile and resilient operator response to novel situations.

Unfortunately, current approaches to behavior authoring remain complex, slow, and limited. Traditional scripts and rule-based methods are generally static/predictable and often complex (Brockington & Darrah, 2002). Scripts often contain weaknesses easily exploited by human players to defeat opponents (Nareyek, 2000). Additionally, due to their non-dynamic static nature, these models often cannot deal with unpredictable human-player tactics, nor can they evolve to handle new environmental situations, or scale up in difficulty to meet, or challenge an opponent's capabilities. These common problems are also associated with state-of-the-art game AI, (Buro, 2003; Spronck, Sprinkhuizen-Kuyper, & Postma, 2004) and hamper the training value and use of those models as well. In order to train to current threat situations, any change in adversary combat tactics or civilian attitudes and patterns of behavior requires model modification by an expert, rendering the use of models as training and research tools expensive and slow in addressing training needs based on changing mission needs. As a result the turnaround time between recognizing a new training need and being able to elicit knowledge and skill development from an operator, based on new behavioral driven models is often delayed and insufficient to support today's rapidly changing asymmetric warfare and the corresponding rapidly evolving threat profiles.

Model Validity/Realism

Pew and Mavor (1998) state that current Semi-automated forces (SAF) and CGF technologies have a significant role in military Modeling and Simulation (M&S); but that highly scripted technologies are limited in use and can be weak. Therefore, they state that observable behaviors in CGFs need to be realistically represented in training and decision aid applications so users can be confident in the plausibility of the models/agents being used. Users tire of CGFs that are predictable (Funge, 1999; 2004) but perhaps what may be worse is that users may learn to "game" the simulation to get the upper hand or 'modify' a trainable model through design or tactics to make it to work in a way they perceive to be more plausible but that may in fact hold little to no psychological, social, or tactical ecological validity. In the end acquiring inappropriate skills and habits creating negative transfer from training to operations that could adversely affect performance in the field. The true benefits of SAFs and CGFs, beyond being simple adversaries, will only be realized if their actions/ reactions are plausible in response to novel external stimuli and internal states.

CURRENT APPROACH

Since behavioral models have rarely been used in real world experimentation in relation to real-time metric integration and adaptive training environments, the team intends to define and implement a theoretical framework to benchmark the evaluation of behavioral model use as training tools; specifically in adaptive constructive scenarios during training research and rehearsal. The primary objective is to develop a set of scientifically-sound evaluation metrics and techniques to assess the modeling architectures' capability, adaptability, and ability to capitalize on and use virtual and real world performance data to rapidly generate models of behaviors for potential use in Distributed Mission Operations (DMO) and LVC training rehearsal environments. These evaluation metrics and techniques will both strengthen the potential of increasing the validity of human behavioral models within a synthetic environment, and enhance the impact of the theoretical advancements in the field of behavioral modeling. The purpose of this activity is to lay a foundation for future training research experiments using behavioral modeling and intelligent constructive adaptive architectures/agents to enhance the elicitation of individual and team performance while training to accelerate the development of adaptive knowledge and skills in warfighters.

Verification and Validation Evaluation Process

The Training Research Team has reviewed different evaluation methods/processes to determine which would provide a strong method/process to assess the use of behavioral/ computational models in training research. Currently the rigorous verification, validation, and accreditation (VV&A) engineering-type process used in software development seems to be the closest method to facilitate and foster the integration and use of models in the training research environment. However, for research purposes, the team plans to adapt a generalized version of the V&V process based on modified successful and accepted model integration and interoperability methods from the following: Mittal & Zeigler, 2008; Mittal and Zeigler, 2007; Wang, Tulk, Wang, 2009; Glenn, Neville, Stokes, Ryder, 2004; Office of Aerospace Studies, Air Force Material Command, 2005.

Phase 1 Model Requirements

In order to properly execute the development and use of valid models, clear specifications and requirements for the use, purpose, and function of the models and a clear definition of the metrics that will determine the usefulness of the models in question must be developed. Essentially, the purpose of a model drives the validation criteria of that model. Models are not expected to be validated in general but to be validated for a specific purpose. In the case of cognitive/computational modeling, validation examines representational accuracy of the conceptual model and, as follows, the executable model in context, and the results of the model in use. Validation has two main components: 1) structural validation, which includes an internal examination of the assumptions, architecture, and algorithms in the context of intended use; and 2) output validation, which determines how well the M&S results compare with the perceived "real world."

The verification process is used to determine if a model/functional requirement/s actually meet the need of a specific well-defined function or purpose. When using models for measurement, as agents or in simulation and training systems, the verification examines transformational accuracy through the model development process; from concept through to the actual use in context. Simply expressed this is the transformation of requirements (functional need) into a conceptual model (descriptive functional properties), then into an executable actionable model (executable functional properties). The verification process can be similar to that employed in general software engineering. (See Figure 1).

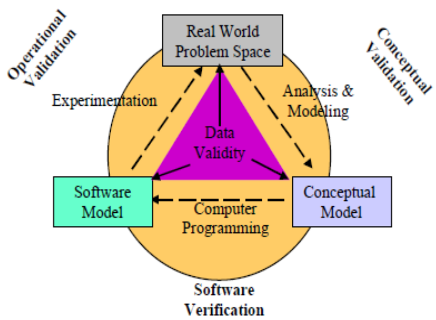

Figure 1. Essential Techniques for Military Modeling and Simulation. From Smith (1998).

In this phase, the research team will define the requirements of two types of threat air platforms that are currently programmed in the threat generation system. Three scenarios scripts using three common adversarial maneuvers will be developed by SMEs. These scripts/ maneuvers will be documented in detail. Then, following the scripts, blue force pilots in high fidelity cockpits will fly against the adversary CGFs; these maneuvers will be recorded in Distributed Interactive Simulation (DIS) protocols using the threat generation system and a control station. Objective measurements such as time, space, and positional information (TSPI) will be collected using the Performance Evaluation Tracking System (PETS). PETS application is designed as a modular, multi-threaded application, capable of robustly handling high volumes of networked entities via various simulation protocols (e.g., DIS, High Level Architecture (HLA)) and calculate measurements at the team, inter-team (package), and teams-of-teams (force) levels. PETS is able to evaluate overall mission performance on all entities within an LVC environment, allowing the trainer to assess the entire picture from both the friendly and enemy perspective. The data can be collected up to 20 cycles per second and consist of simple TSPI to more complex calculations. PETS has the capability to add custom entity modules that can handle custom information packets "passed through" the network (Watz, Keck, & Schreiber, 2004; Portrey, Keck, and Schreiber, 2005; Schreiber, Stock & Bennett, 2006).

DIS data will also be recorded and collected from the manned threat simulation system as experienced adversary pilots fly out the scripts against the blue force pilots via a distributed network. This information will be used to test and evaluate the validity of the models, the viability of the modeling architectures for using in training research and the model's capability to utilize objective human/team/constructive agent performance and simulation data for the purposes so defined.

Phase 2 Metric Development

As the process is developed, the way comparisons are made needs to be taken into consideration. In a strict V&V environment, a validation metric is required when the original requirement is written for the test phase. Such metrics are typically objective and may include but are not limited to a percent pass/fail threshold and subsequent mark when evaluating the difference between the results of a model and the referent. Often in evaluating physics-based models there is a dependency on mathematical representations of the problem space which are relatively straightforward. They usually represent a comprehensible quantitative metric; a yes/no question and because the physical sciences are mature, an experimental criteria with clear referents; equally powerful and unlikely to be disputed.

However, in the cognitive and social science modeling realm, metric strength can be lost; lost in the process of translation, because determining causation by tracing particular event outcomes, in a one-to-one fashion to the original causal factor or set of culminating events, can be less than a straightforward process. In addition, determining the level of uncertainty of causation in social and cognitive models is probably the single greatest challenge during validation. More often than not, the more complex a model and/or simulation environment is, the greater the uncertainty factor can be; leading to uncertainty playing a critical role in the evaluation of the model.

The empirical focuses on a taxonomy of observable actions to facilitate automated human performance data collection (e.g., keystroke actions, voice communications, etc) of data that are conceptually relevant to the model evaluation process. From the model's point-of-view the methodology identifies events attributable and mappable to the observable actions and the relevant associated evaluation criteria. Table 1 represents a summary of candidate subjective measures that are being considered when conducting an evaluation of models for incorporation and practical use in Training Research Environments.

Table 1 Candidate Subjective Measures

Architecture
Architecture stability
Available to the public (non-proprietary and modifiable)
Usability (ease of integrating existing performance data into the modeling architecture)
Adaptability (time to develop/update and implement a human behavior into the model)
Network protocol requirements are met (Distributed Interactive Simulation ; DIS, High Level Architecture; HLA)
Supported by sufficient documentation
Model
Does the model in context support ecological validity
How realistic are the behaviors of the model in regards to its intended use
Plausibility of model based behavior
Does the model satisfy the initial purpose and requirement?
To what level is the model agnostic with respect to a Framework or Architecture
Re-usability of the model
Is the use of the model for a given purpose valid (e. g., to measure human skill a psychologically plausible model is required)
Fidelity - model's ability to accurately represent specified behaviors
Can the models behavior be correlated with human behavior if appropriate to do so
Generalizability of the model
Model Cont.
Under what conditions does the model's credibility, believability, and reliability break down
How easily can different/new human related data be injected (is the model highly data base structure dependant)
Is the model able to generate believable CGF
Is the model's behavior transparent to users not involved in the core modeling process

Note. Summary of Candidate Subjective Measures modified from By Order of the Secretary of the Air Force, 1996; Petty 2010; Department of the Army, 1999.

In regards to objective performance measurements from PETS, these data will serve three-fold. This data will be used to feed the model with the information needed to make the model/s more realistic. PETS data will also be used as a metric to observe how fast the architecture can modify/adapt the model to address new behaviors. They will also be used to evaluate how accurately the model's PETS measurable behavior correlates to the human PETS measureable behavior.

Phase 3 Process Documentation

Some recommendations made by Tollefson, Martin, and Fletcher (2008) for the modeling processes include a complete conceptual description of the behavior models. Development and documentation of standards; standards followed throughout the development, test and use process. Implementation of these standards would facilitate traceability from the real-world behavior to the final execution of the model. This can be done by ensuring a clear link between knowledge acquisition, engineering processes, and the resulting model. It will be important to define and document model assumptions, capabilities, limitations, and risk/impact within each context of purpose for which the models is expected to perform. Such a living document should also describe model strengths, weaknesses, categorical usefulness, or limitations, data type requirements, and perhaps such properties as interoperability, usability, re-usability and reliability. Any documentation should also include any recommended remedial actions.

FUTURE DIRECTIONS

So the current state of affairs, with respect to human behavior modeling, and automated real-time metric collection in the development of constructive models as well as the need for dynamic adaptive systems for training is ripe. However, along with a leap in the technological application and potential hybridization of psychologically plausible cognitive architectures and agent executive components comes a whole host of methodological and technological as well as pragmatic concerns about the use of these systems in training, especially in LVC environments. While on the one hand researchers and users would like to get to the point where no one can tell whether it's real or Memorex, researchers still have to be able to retain traceability to the conceptual representation of a human behavior based models or AI based agents expected behavior/functional response when acting in the virtual world or imposing activity on the real world; maintaining the ability to trace causality of behavior. Mainly due to the fact that there is a potential for resulting behavior to come from an emergent process; an entity or system behavior which by definition cannot be broken down or traced to the sum of its parts. So, one core approach is to retain modularity in behavioral components, to retain sequential behavior tractability, and, if possible, to also retain the ability to take a reductionist testing approach in complex applied environments.

The research team views this process and the potential outcomes as very exploratory and basic in nature. However, the venue in which to integrate these processes, to further research in training, is very dynamic and complex so the team is driven to formulate and use some formal or semi- formal method of evaluation and integration of models for use in behavior modeling or as AI agents before supporting future use of models in these types of applied training and research environments. As more information is gathered on the feasibility, validity, and reliability of using adaptive models for training research, and as successful modeling architectures are identified, the team plans to continue their venture into rapid modeling development assessment by tackling the questions listed in Table 2.

Table 2 Future Candidate Measures

Candidate Measures
Can the models output be traced to input to determine a logical chain of causality
Can the model generate tactically believable adversarial forces
Is the model compatible with other models with respect to interoperability and sharing functional capacities when required to do so
Can the model perform without interferences in a federation of models
What is the model's level of domain independence
Can the model support data acquisition and understanding of an individual's intent, behavior, or response
Can the model pass a Wizard of Oz type test by replacing AI with Human SMEs and comparing results
Does the model support, in a tutoring fashion, adaptive training
Does the model support or exploit automated measures capture
Does the model support real-time or near real-time data analysis
Does the model execute critical mission actions accurately
Can the behavior model/AI agent when used for an intended purpose positively affect training outcomes
Can the use of models in adaptive training systems positively affect the outcome of training

Note. Summary of Future Candidate Measures modified from By Order of the Secretary of the Air Force, 1996; Petty 2010; Department of the Army, 1999.

SUMMARY

The research team hopes to support the development of a process the larger community will equally contribute to and when the dust settles embrace for whatever value the results of such an effort may hold. When working on the edge of new vistas in science and engineering, rarely does one individual or team discover truth in theory or solidify methods and unifying processes that support the discovery of ground truth and confidently determine what the new findings can and should be used for. It is usually only through collaboration and/or our competitive nature that scientific/engineering communities break new ground and gain stability in process, methodology, and use of the new tools. The team recognizes that currently in the realm of this inquiry, there may be little if any agreed upon unifying theories, methods, or validation approaches for what is currently being attempted. So in nearly every way we recognize this effort to be an exploratory process. But only by way of trying to begin to define and construct quality methods and processes can researchers generate viable systems and theory, and sustainable Cognitive/Computer scientific and engineering processes in such a complex applied environment while attempting to with sufficient level of rigor also implement a reductionist approach during inquiries in these areas of research and application.

If these recognized shortcomings are overcome to any extent the state of the art for the use of behavior models in building realistic threat generation and adaptive training systems will have achieved an evolutionary leap in practice through a technology, application, and integration process paradigm shift. Such progresses should afford the use of more human behavior type and/or hybrid AI type models in not only training environments, but in adaptive constructive agent systems, automation, and Predictive Threat Modeling.

REFERENCES

Anderson, J. R., & Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah, NJ: Erlbaum.

Best, B. J., Lebiere, C., & Scarpinatto, C. (2002). A model synthetic opponents in MOUT training simulations using the ACT-R cognitive architecture. In *Proceedings of the Eleventh Conference on Computer Generated Forces and Behavior Representation*. Orlando, FL: National Security Industrial Association.

Brockington, M., & Darrah, M. (2002). How not to implement a basic scripting language. In S. Rabin, (Ed.), *AI game programming wisdom* (pp. 548-554). Boston, MA: Charles River Media.

Buro, M. (2003). Real-time strategy games: A new AI research challenge. In *2003 Proceedings of the International Joint Conference on Artificial Intelligence* (IJCAI) (pp. 1534-1535). Acapulco, Mexico: Morgan Kaufmann.

By Order of the Secretary of the Air Force. (1996). Air Force Instruction 16-1001: Operations Support: Verification, Validation, and Accreditation (VV&A), 1 June 1996. Washington, DC: Department of the Air Force.

Department of the Army. (1999). *Department of the Army Pamphlet 5-11: Verification, Validation and Accreditation of Army Models and Simulations* (30 September 1999). Washington, DC: Author.

Freeman, W. D., & Burns, W. R. Jr. (2010). *Developing an Adaptability Training Strategy and Policy for the Department of Defense (DOD)* (ADA531765). Alexandria, VA: Institute for Defense Analysis.

Fu, D., & Houlette, R. (2004). The Ultimate Guide to FSMs in Games. In S. Rabin (Ed.), *AI Game Programming Wisdom 2, 2003* (pp. 283–302). Boston, MA: Charles River Media.

Funge, J. D. (1999). *AI for Games and Animation: A cognitive modeling approach*. Natick, MA: A. K. Peters.

Funge, J. D. (2004). *Artificial Intelligence for Computer Games: An Introduction*. Natick, MA: A. K. Peters.

Glenn, F., Neville, K., Stokes, J., & Ryder, J. (2004). Validation and Calibration of Human Performance Models to Support Simulation based Acquisition. In *WSC '04 Proceedings of the 36th conference on Winter simulation (*pp. 464-471). IEEE.

Grisogono, A. (2010). *TTCP Technical Report, JSA Action Group 14 Complex Adaptive Systems for Defence: Synthesis Report: Conceptual Framework for Adaptation*, May 2010 (DOC – AG14 - CFA #3-2010). Paper presented at the Operational Adaptation Conference. The University of Edinburgh, 22-24 June 2010.

Hoang, H., Lee-Urban, S., & Munoz-Avila, H. (2005). *Hierarchical plan representations for encoding strategic game AI*. Paper presented at the Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE-05), Marina del Rey, CA.

Klir, G. J. (2003). An update on generalized information theory. In *Proceedings of ISIPTA '03*, (pp. 321–334). Lugano, Switzerland: Carleton Scientific.

Lamoureux, T. M., Bandali, L. M., Martin, B., & Li, Z. (2006). *Team modeling: Review of experimental scenarios and computational models* (DRDC Toronto No. CR2006-092). Guelph, ON: Human*systems*® Incorporated. PWGSC Contract No. W7711-047911/001/TOR Call-up No. 7911-05.

Mittal, S., & B. P. Zeigler. (2008). *DEVS unified process for integrated development and testing of system of systems.* Presented at AFCEA-GMU C4I Center Symposium: Critical Issues In C4I**.**

Mittal, S., Zeigler, B. P., Martín, J. L. R., Sahin, F., & Jamshidi, M. (2007). Modeling and simulation for systems of systems engineering. In *Systems of systems engineering* (Chapter 5). Wiley Online Library. John Wiley & Sons, Ltd.

Mueller-Hanson, R. A., Nelson, J. K., & Swartout, E. (July 2009). *Proof of Concept Research for Developing Adaptive Performance: Task 2 Report, Validation Plan*. Arlington, VA: PDRI.

Nareyek, A. (2000). Review: Intelligent agents for computer games. In T. A. Marsland & I. Frank (Eds.), In *Revised Papers from the Second International Conference on Computers and*

Games (CG '00) (pp. 414-422). London, UK: Springer-Verlag.

Newell, A. (1994). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Office of Aerospace Studies Air Force Material Command. (September, 2005). *Analysis of Alternatives Handbook.* Kirtland AFB, NM: Department of Defense.

Office of the Under Secretary of Defense (Personnel & Readiness). (September, 2010). *Strategic Plan for the Next Generation of Training for the Department of Defense*. Washington, DC: Department of Defense.

Orkin, J. (2004). Symbolic representation of game world state: Toward real-time planning in games. Paper presented at the Challenges in Game AI: Papers of the AAAI'04 Workshop, San Jose, CA.

Petty, M. D. (2010). Verification, validation, and accreditation. In J. A. Sokolowski & C. M. Banks (Eds.), *Modeling and simulation fundamentals theoretical underpinnings and practical domains.* Wiley Online Library. John Wiley & Sons, Ltd.

Portrey, A. M., Keck, L. M., & Schreiber, B. T. (2005). Challenges in Developing a Performance Measurement System for the Global Virtual Environment. *In Proceedings of the 2005 Spring Simulation Interoperability Workshop*. San Diego, CA: SISO, Inc.

Pew, R. W., & Mavor, A. S. (Eds.). (1998). *Modeling human and organizational behavior: Application to military simulations*. Washington, DC: National Academy Press.

Randolph, W., & D. Sagan. (2003). OneSAF uses a repeatable knowledge acquisition process. *In Proceedings of 2003 Industry/Interservice Training Systems Conference*. Orlando, FL: National Security Industrial Association.

Schreiber, B. T., Stock W. A., & Bennett, W. Jr. (2006*). Distributed Mission Operations within-simulator training effectiveness baseline study: Metric development and objectively quantifying the degree of learning.* (AFRL-HE-AZ-TR-2006-0015-Vol II). Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Smith, R. D. (1998). Essential techniques for military modeling & simulation. In *Proceedings of the 30th conference on Winter simulation* (WSC '98). Los Alamitos, CA: IEEE Computer Society Press.

Spronck, P. H. M., Ponsen, M., Sprinkhuizen-Kuyper, I. G., & Postma, O. E. (2006). Adaptive game AI with dynamic scripting. *Machine Learning, 63*(3), 217-248.

Spronck, P. H. M., Sprinkhuizen-Kuyper, I. G., & Postma, E. O. (2004). Difficulty scaling of game AI. In A. El Rhalibi & D. Van Welden, (Eds.), In *GAME-ON 2004 5th International Conference on Intelligent Games and Simulation* (pp. 33–37). Belgium: EUROSIS.

Tollefson, E. S., Martin, M. S., & Fletcher, A. (2008). *OneSAF Objective System (OOS) Behavior Model Verification* (TRAC-M-TM-08-05315, August 2008). Monterey, CA: TRADOC Analysis Center.

Tollefson, E. S., Schamburg, J. B., & Yamauchi, H. M. (2006). A methodology for conducting composite behavior model verification in a combat simulation. In L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, & R. M. Fujimoto, (Eds.), *Proceedings of the 2006 Winter Simulation Conference* (WSC 2006) (pp. 286-1293) Monterey, CA, December 3-6, 2006.

United States Air Force. (November 2004). *United States Distributed Mission Operations Implementation Plan.* Washington DC: Deputy Chief of Staff and Space Operations.

Wang, W. G., Tolk, A., & Wang, W. P. (2009). The levels of conceptual interoperability model: Applying systems engineering principles to M&S. In *Proceedings of the Spring Simulation Multiconference* (SpringSim '09). San Diego, CA: Society for Computer Simulation International.

Watz, E., Keck, L., & Schreiber, B. T. (2004). Using PETS Software to Capture Complex Objective Measurement Data from Distributed Mission Operations (DMO) Environments. *In Proceedings of the 2004 Spring Simulation Interoperability Workshop.* Arlington, VA. Paper available as 04S-SIW-143.

Wray, R. E., Laird, J. E., Nuxoll, A., Stokes, D., & Kerfoot, A. (2005). Synthetic adversaries for urban combat training. *AI Magazine, 26*(3), 82-92.

Zacharias, G. L., Macmillan, J., & Van Hemel, S. B. (Eds.). (2008). *Behavioral modeling and simulation: From individual to societies*. Washington, DC: The National Academic Press.