Validating Human Behavioral Models for Combat Simulations Using Techniques for the Evaluation of Human Performance

Simon R. Goerger
Naval Postgraduate School
MOVES Institute
Monterey, CA 93943-5118
831-656-3733
srgoerge@nps.navy.mil

Keywords: Validation, Human Behavior, Cognitive Task Analysis

Abstract

As Department of Defense (DoD) leaders rely more on modeling and simulation to provide information on which to base strategic and tactical decisions, simulation credibility becomes more important. Prior to their use in simulations and analytical studies, DoD models are required to undergo the verification, validation, and accreditation (VV&A) process in an attempt to establish an acceptable level of credibility. In general, the human behavioral model validation process, as outlined by the Defense Modeling and Simulation Office (DMSO), is not extendable to meet requirements for validating the varied and complex behavioral models in use or under development for DoD simulations. This paper reviews several issues with validating human behavior representation (HBR) and identifies potential practices for enhancing the validation process for current and future human behavioral models for use in or application to combat simulations.

INTRODUCTION

Developing a cognitive model to operate computer generated forces is difficult at best. Ensuring it adequately represents the human behavior it is designed to emulate in the multitude of nonlinear environments it is asked to perform in is nearly impossible. However, if the Department of Defense is to use models and simulations to support training and testing, "it is not only sensible, but it is also the law" [1] to verify, validate, and accredit these models.

"In the military context, the most highly validated models are physiological models and a few specific weapons models. Few individual combatant or unit-level models in the military context have been validated using statistical comparisons for prediction; in fact, many have only been grounded.[1] Validation, clearly a critical issue, is necessary if simulations are to be the basis for training or policy making." [2]

With physics based models, there are established procedures for performing VV&A that allow developers and users to understand the strengths and limitations of a model. For cognitive models, the procedures are not as well established and are often limited in their execution and in the information they provide. Understanding the human thought and decision making processes is complex and evolving. Thus, developing a theoretical model and implementing it in code is problematic. This adds to the difficulty of gaining credibility for a model.

Verifying code, validating the performance of a model, and accrediting a model for use in a simulation are the three aspects for gaining credibility for a model. All three aspects of official certification are important, but the scope of this paper does not allow sufficient space to address issues with all three phases; this paper focuses on *validation* of human behavior representation model implementations.

The remainder of this paper covers the background behind model validation and discusses issues with the current process of validating a cognitive model before moving to the presentation of three potential techniques to address these shortcomings. Conclusions follow along with an outline of proposed future work to explore the proposed techniques.

BACKGROUND

Whether or not a model needs to go through the VV&A process is often in debate. As of 1994, the DoD Directive (DoDD) 5000.59 states: "M&S applications used to support the major

[1] Grounding is a form of face validation which demonstrates "that simplifications do not detract from (the) credibility" of a model. [2]

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **2004**	2. REPORT TYPE	3. DATES COVERED **00-00-2004 to 00-00-2004**

4. TITLE AND SUBTITLE **Validating Human Behavioral Models for Combat Simulations Using Techniques for the Evaluation of Human Performance**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Postgraduate School,Moves Institute,Monterey,CA,93943-5118**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**11**	

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

DoD decision making organizations and processes shall be accredited[2] for that use by the DoD Component for its own forces and capabilities." [3] DoD Instruction 5000.61 expands the list of models which require accreditation to include any model used for joint training or exercises as well as any model or simulation for which the DoD Component deems accreditation is warranted. [4]

As one of the three phases of the VV&A process, the purpose of validation is to determine if a model adequately replicates the real world action/behavior it was intended to represent. The validation process for any model is performed by a validation agent assigned or hired by the individual or agency responsible for the overall accreditation process of the model; [5] normally, the validation agent is the model sponsor. [6] To facilitate the accreditation process, the validation process should begin when a model is first being conceptualized and continued until model modifications and usage are complete.

The remainder of this background section will cover more specifics for validating models replicating human behavior. These models are often referred to as human behavior representation (HBR) models.

Validation Process for Human Behavior Representations (HBR)

A validating agent seeks to determine how well the human behavior model results match system requirements and a referent. Based on the Defense Modeling and Simulation Office's Recommended Practices Guide (RPG) this is accomplished at four distinct phases of model development. These are: 1) the design of the conceptual model; 2) the generation of the knowledge base; 3) the implementation of the model and its knowledge base; and 4) the integration of the model into the simulation. [7] The amount of credibility a model initially has is often based on how well the validation agent believes the model performed during each of these phases.

DMSO's "Validation of Human Behavior Representation," outlines seven high-level tasks for validating an HBR. These seven tasks are:

a) Identify system requirements and acceptable conditions for a potential HBR model
b) Collect referent to assess correctness of HBR performance
c) Validate the HBR's conceptual model using human performance referent and requirements
d) Dissect the HBR's conceptual model to identify complex areas of the model to focus future validation activities (focusing on results validation)
e) Validate the HBR's knowledge base using human performance referent and requirements
f) Dissect the HBR's knowledge base to identify complex areas of the model to focus future validation activities
g) Validate the integrated HBR model using human performance referent and requirements directed toward the most complex areas of the model as identified by the complexity analysis of the conceptual model and knowledge base. [7]

Figure 1. shows where these tasks would lay in the VV&A process for a cognitive model.

Referent

As the codified body of knowledge, referent for HBR is normally collected from one or more resources. [8] Many of these resources are validated models. Examples are models of specific aspects of human behavior, sociological phenomena, and the physiological processes underlying human behavior. Referent is also collected from validated simulations of human behavior (live, virtual, or constructive), empirical observations of actual operations, experimental data, and subject matter experts (SMEs). [7]

The "Key Concepts of VV&A" document describes six categories of correspondence, or the agreement of a model to different levels of abstraction, usable for determining referent for HBR: computational, domain, physical, physiological, psychological, and sociological. [9] This paper will define three of these which were used in the National Research Council study conducted in 1998 and published in Modeling Human and Organization Behavior. [2] These categories are domain, physiological, and psychological correspondence.

[2] As defined by DoDD 5000.59, accreditation is the "official certification that a model or simulation is acceptable for use for a specific purpose." [3]

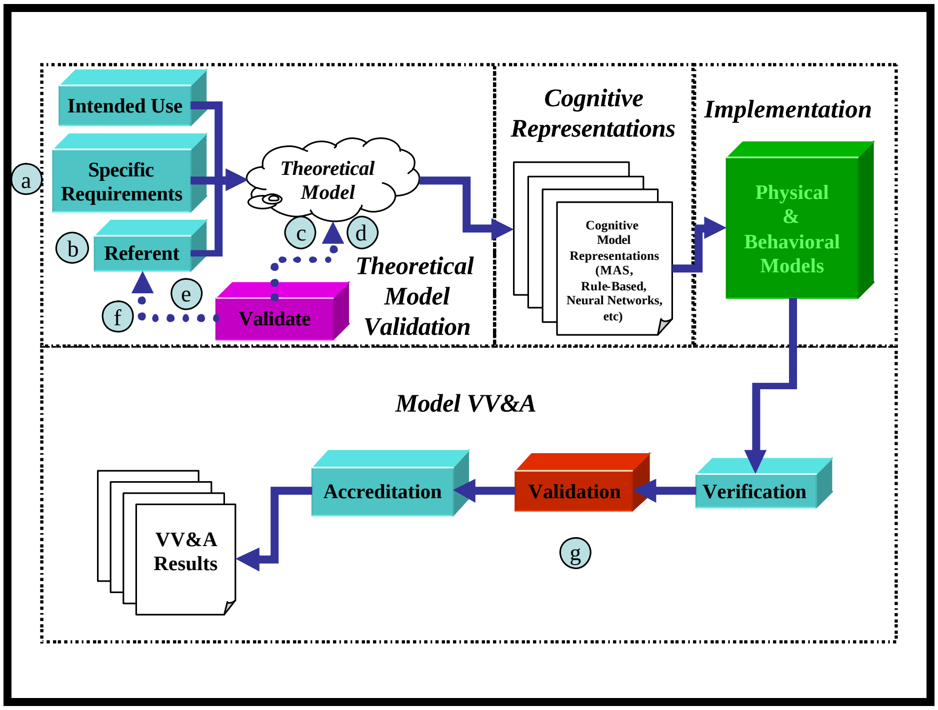

Figure 1. Verification, validation, and accreditation tasks for a cognitive model

Domain Correspondence

Domain correspondence addresses the use of SMEs to examine the knowledge base and outcomes of human behavior in their respective areas of interest. The data collected is normally qualitative in nature and leads to referent viable for face validation, a form of validation often equated to a Turing Test. [7]

Physiological Correspondence

Physiological correspondence resembles many of the techniques used to validate physical models. It uses information from neurologists, neurosurgeons, and/or physiologists to determine if a model's components react similarly to the portion of the brain they are asked to replicate. This form of validation has become more viable over the last two decades due to physiological advances in the understanding of the human nervous system. Physiological correspondence is considered by some an immature area of study but has demonstrated use in validating neural networks. [7]

Psychological Correspondence

The SME for psychological correspondence is the psychology professional. Similar to the use of SMEs in domain correspondence, the psychologist can provide a qualitative analysis comparing real world behavior to model results to determine if the model exhibits human like behaviors. Psychological correspondence can also be cultivated from the numerous volumes of experimental data on human performance in varying real world scenarios. [7]

Behavior Model Representations

Over the past forty years, model developers in the artificial intelligence (AI) and artificial life (AL) communities have used numerous techniques to implement their theoretical models of human behavior. What follows is a short description of five behavior model representations: Rule-Based, Bayesian-Network, Neural-Network, Agent-Based, and Multi-Agent System.

Rule-Based

A rule-based or knowledge based system endeavors to imitate human behavior using an enumeration of steps with causal if/then association "using rules represented as symbolic expressions". [10; 11] The representation requires a comprehensive identification and coding of possible situations an agent or entity may encounter and resulting viable actions for those conditions. SMEs are routinely used to identify probable and possible situations. These conditions and associated actions must be entered into the knowledge database for the model. Problems arise with rule-based models when a situation occurs that is not represented in the model's database. Such situations can result in model failure, inappropriate action(s), or the construction of new rules to deal with the current state of the simulation.

Bayesian-Network

A Bayesian-network or belief network represents the dependencies between variables to provide a succinct design of a joint probability distribution. The network is a directed graph where nodes are sets of random variables; directed links connect node pairs signifying which nodes have a direct effect on other nodes; each node has a conditional probability table representing the quantifiable impact each parent node has on the child node's value; and the graph has no directed sequences determining the specific path to be taken or result. The links, representing direct conditional dependency between nodes, and the probability coupled with each link are typically established by SMEs. Uncertainties can be applied to each node to help make runs stochastic. A deterministic run is executed when a child node's values are derived exclusively from the inputs of the node's parent(s).

A Bayesian-network can reason from effects to causes (diagnostic inference), from causes to effects (causal inference), between causes of a common effect (intercausal inference), or by combining two or more of the above (mixed inference). One of the obstacles with producing a Bayesian-network comes from the inability of SMEs to ascertain all the nodes and directed links essential for an implementation in a particular domain. Finally, determining the probability weights for each link is often considered the most complex phase of creating and modifying a Bayesian-network. [11]

Neural-Network

A neural-network is analogous to a Bayesian-network. However, a neural-network is a cognitive model representation that endeavors to duplicate some of the properties of the human brain instead of replicating the dependencies between variables. It consists of numerous simple components (neurons) operating in parallel with no central control. The connections (arcs) between nodes have weights. These weights are adjusted by the system during the model's training phase based on a series of training inputs and expected results.

Once the model is trained to generate the appropriate results for the given inputs, the network produces outcomes based on the nature of the interaction of the internal network of nodes and the connection topology. [11] This cognitive representation is often used when there is a limited set of inputs and possible outputs. Neural-networks have been used to successfully analyze handwriting on letters to identify zip codes.

Neural-networks are often associated with expert systems that recognize complex data sets and produce rational behavior. These systems attempt to imitate reasonable behavior for procedural or reactive tasks. Due to the complex nature of the network's interconnected nodes, it is very difficult to perform more than a face evaluation for even the most simplistic behavioral model coded as a neural-network. Thus, neural-networks are frequently regarded as "black box" AI implementations. [11]

Agent-Based

Agent-based technology affords an ability to exhibit intelligence through computer simulated objects that can identify characteristics of the environment, real world or simulated, and then act on those observations. [11] There are several types of agents; two of these are reactive and rational agents. Agents have intent which guides their response to their perceived environment. A reactive agent uses the last set of sensory inputs to determine which action(s) to execute. Often a condition-action rule (this is the state of my perceived world, this is the action I take) is used for these agents. A rational agent also uses sensors to observe its environment then performs actions on the environment using effectors. However, unlike reactive agents, rational agents maintain a state of situational awareness based on their previous knowledge of the world and current sensory inputs. [11]

Multi-Agent System (MAS)

A multi-agent system (MAS) is a behavior model representation with autonomous or semi-autonomous software agents that produce adaptive and emergent behaviors. Adaptive behavior is the process of fitting oneself to the environment and emergent behavior is generated at a higher cognitive level based on the behaviors and interactions of agents at a lower level. [12] The MAS model uses a bottom-up approach where software agents make independent micro-decisions that generate group level macro-behaviors demonstrating emergent behavior. An MAS can use any form of agent-based software technology (reactive, rational, goal-based, utility-based, etc.) with agents described as possessing intentions that influence their actions.

Multi-agent systems are used in relatively large domains where non-linearity is present. [13] The MAS, limited only by the physical constraints of the simulation boundaries, uses an indirect approach to search the large domain for viable results. Another feature of a multi-agent system is its capacity for agents to evolve over time to create new agents which are normally more adept at surviving/thriving in the virtual environment. [14] Some MAS agents have been coded with a "brain lid" to allow inspection of the agent to determine its situational awareness and the decision processes it used to select a specific action. [15] Interrogating such agents allows one to potentially view the reasoning behind the actions of the agent. [16]

Subject Matter Experts (SMEs)

Besides identifying requirements and collecting referent, SMEs are used to perform validation. In fact, to date, the most common means of validating cognitive (HBR) models has been through face validation using SMEs. [8] Often this technique uses an SME to exercise the HBR in a scenario where the SME manipulates the model through the simulation space by issuing orders or varying stimulants, observing resulting behavior, and determining whether the observed overt behavior meets a user's requirements for realism. This is often done using qualitative referent. [7]

SMEs come from many realms based on the validation needs and model's intended purpose. SMEs are normally selected by the validation agent or are assigned by independent agencies. Their selection is often based on availability, expertise, familiarity with simulations, the focus of the validation effort, and the type of validation techniques being utilized. Occasionally, SMEs will receive additional training and or certification prior to beginning their validation effort; however this is neither a requirement nor a routine practice by validation agents.

ISSUES

To date, formal validation is not always attainable. "Current state-of-the-art proof of correctness techniques are simply not capable of being applied to even a reasonably complex simulation model. However, formal techniques serve as the foundation for other V&V techniques." [17] In the validation of cognitive models there are many issues which make it difficult to accomplish and even harder to ensure uniform standards of implementation. This section will outline four areas identified by DMSO and address five other factors: referent bias, the use of SMEs, model representation, the limitations of validating cognitive models using face validation of overt behaviors, and cost.

DMSO Validation Issues

DMSO has identified four factors that make validation of cognitive (HBR) models difficult. The first is that for even simple human behaviors, the set of possible actions is normally very large. This makes it difficult to ensure examination of all viable solutions. The next issue is the general non-linear characteristic of the constrained space of consideration. The non-linearity of the space prevents a simple causal relationship to be drawn between situational parameters and resulting actions. Third is the propensity of some behavioral models that introduce stochastic features to their models to allow the model to exhibit unpredictability. This "unpredictable" characteristic, unless it can be forced to be deterministic, often makes repeatability impossible for a model therefore making model validation more difficult, and frequently impossible. The final obstacle to validation identified by DMSO is the chaotic behavior exhibited by behavior models that are sensitive to initial and boundary conditions. Models with such issues are limited to the breadth of their validation and to the set of scenarios where they exhibit stable behavior. [7]

Referent

Using the three most common formats of referent as outlined in <u>Modeling Human and Organization Behavior</u> helps identify some of the issues with validating cogitative models. [2] Reviewing cognitive models and types of correspondence used for their validation reveals some of the difficulties in obtaining and using referent. As stated previously, domain and psychological correspondence gather their referent from SMEs. Thus, these two forms of correspondence are generally subject to SME bias which often limits their use to providing qualitative data, and routinely results in face validation of the model. Because of the vast spectrum of potential situations and human responses, the identification and collection of referent are often limited. This reduces the available pool of data available to evaluate the capabilities of a model and limits the number of situation a model can be specifically tested for compliance.

The validation process is inconsistently applied because it is performed by multiple V&V agencies with non-standard criteria or non-uniform referent. [7] This often leads to an invalid comparison of cognitive models due to the non-uniform means of validation. The difficulty in collecting referent for each category of correspondence for use in validating cognitive models for different domains is an issue. Human performance data is an area in which numerous resources have been provided to collect referent. Models validated using more than one category of correspondence often focus on domain and psychological correspondence, but routinely limit face validation of overt behaviors.

All validation techniques have limitations. There are two significant limitations of the HBR correspondence described above. The first deals with the unrealistic requirement of domain correspondence to search very large and nonlinear behavior spaces. The second concerns testing for psychological and physiological correspondences. These two forms of correspondence usually require the use of extensively validated models of psychological and physiological phenomena to produce referent. [18] In essence, one must find results from other valid HBR models or build and validate another cognitive model to provide referent for validation of a new cognitive model. This dependence on other models makes validation using psychological and physiological correspondences tenuous at best.

Model Representation

Face validation addresses the overt behaviors of a model. These behaviors are the results of model computations and allow validation agents to correlate inputs with outputs and compare them with referent for accuracy. The validation technique is routinely used because all functioning computerized models take some set of inputs and produce results. Problems occur when a model is fed unique/new inputs for which real world outputs have not been recorded. In these situations, it is not clear if the model's results adequately represent probable or possible actions.

Each model representation has limitations to the type and amount of data it can make available to assist in the validation process. As stated earlier, understanding the relationship between nodes of a neural-network and the impact each node has on the final results is a complex matter at best. However, an MAS implementation may provide access to the information known or considered by each entity and the impact each piece of information has on the model's determination of its results. The different class of data accessible by each model constrains the style of validation techniques available for use to validate an HBR model.

Because of the diverse nature of human performance and the non-linear, chaotic relationship between environmental conditions and human actions, merely looking at the overt behaviors of a model limits the level of confidence one can have that the model will replicate reasonably human behavior with even minor modifications to environmental inputs. Being able to access the underlying implementation of a model to view its situational awareness and algorithmic thought process will likely provide a more accurate evaluation of the model's ability to exhibit human behavior over a broader range of environmental conditions and missions.

Subject Matter Experts

The use of SMEs to evaluate the results of a simulation is analogous to the use of introspection. In the 1920s, behavioral psychologists discounted introspection as a means to have experts explain their actions while they executed tasks. Introspection techniques also used individual reflections to look back on prior situations. This was determined to be problematic as experts often found justification for actions that were instinctive. [19] However, despite the limited use of introspection in psychology, validating agents still use "behavior visualization techniques [which are similar to introspection, because these techniques] can greatly help SMEs examine simulation results, particularly for simulations with which they [the SMEs] can interact in real time." [8]

Validating a model using psychological correspondence has potential issues with the qualitative nature of the referent and unintentional bias of the psychological experts. However, psychological correspondence testing has the potential for greater credibility as more models of emotional phenomena are codified and validated. These validated models may provide baseline data and reduce the need for an exhaustive search of psychological problem space to identify appropriate referent. This holds the most promise for models that incorporate stress and emotion. [7]

According to a meeting of validation experts at Foundations '02, there are at least three major issues with the use of SME: perspective, performance, and perception. [20] Perspective deals with an SME's ability to maintain focus on the intended purpose of the model. According to DMSO RPG, models are to be validated for a specific use. An SME may lose this focus as he allows his experiences with the real world to cloud his view on what the model should have the capability of doing. Performance deals with the SME's ability to execute the validation process. This ability may be hampered by other demands on the SME's time, the availability of data, the ability or desire to comply with specified validation procedures, or the ability of the expert to understand the simulation. Finally, perception addresses the bias an expert brings to the process based on his education, training, real world experiences, exposure to simulations, and organizational loyalties. These factors could color the lenses of the SME's microscope or unduly focuses the search area on certain aspects of a model's performance.

Overt vs Cognitive Analysis

As discussed in model representation, there are many problems with validating HBR models simply on their overt behaviors. As we attempt to expand the functionality of HBR models to operate in open systems where multiple methods and variable situations exist that the model has to operate in, constrained HBR models may better meet the needs of the simulation. [21] However, using results based, overt behavior validation of HBR systems often fails to capture the flexibility of the model. This method of validation also falls short of covering the dynamic problem space in which such a model could be asked to operate. There is a need to understand the underlying cognitive process of the model to allow its potential validation for more than the limited set of situations for which it can be tested. Time and model representation implementation considerations may limit the ability to view and evaluate the cognitive process of the model. The ability to view such cognitive processes holds potential for allowing a wider and more complete review of the model's capabilities in tested and untested circumstances.

Cost

Cost is a general term that can be calculated using various means. In the area of the validation process, these costs are not always well understood nor are the resources easily identifiable. [22] The reality of the situation is that validation is routinely left to the end of the model development process and limited to the remaining funds and time available. [23]

So how does one maximize the level of validation with the available time and funds? One method used by validation agents is to limit the number of SMEs and simulations runs. Depending on the study process, this could result in divergent results from SMEs and an inadequate number of data points to provide statistical significance for the results of the validation effort. This requires validation agents to scrub the qualitative results of SMEs and use other SMEs to referee the conclusions of the studies. The end result is a very narrow validation of the model based on potentially statistically insignificant results of limited qualitative and quantitative data.

POTENTIAL PRACTICES

With issues of cost, SME bias, model constraints, referent collection, and the limitations of face validation of utilizing only inputs and overt behaviors there is a vast field of potential targets on which to focus our efforts for improvement of validation processes. Solving any one of these issues brings us closer to a more complete and meaningful validation. The use of qualitative validation techniques is limited in their application to the analysis of qualitative information. Therefore, we are currently relegated to using SMEs to perform validations of most HBR models. The following subparagraphs address how one might better identify, prepare, and utilize SMEs in the validation of HBR after its integration into a simulation and prior to its utilization in training or analysis.

Subject Matter Experts

The Foundation '02 Special Topics Session on "SME Use in M&S V&V" discussed potential actions which could help to improve the capabilities and use of SME to validate models. Two of these were a set of standards for identifying and accrediting SMEs and training SMEs to help provide them with a set of skills to help them focus their validation efforts. [20]

Selecting and certifying SMEs would ensure a minimum set of standards for SMEs, provide validation agents with a pool of potential SMEs, and increase the credibility of SMEs. It was recommended that the community look to the legal profession guidelines for determining technical experts for potential characteristics of an SME. Although this proposal has some concerns, there has been limited objection to a system that would help establish standards. Requirements for certification should include official education in the area of expertise, practical experience in the area of expertise, and familiarization with models and simulation(s). Additionally, a responsible agency would need to be identified to certify SMEs and maintain the list of certified individuals for each specialty.

Along with certification is the requirement for a training program to ensure potential SMEs can gain the necessary knowledge of models and simulations and the validation process so they can prepare to complete a certification process. This program might also provide refresher training for those who wish to maintain their certification as an SME. To help limit the bias of SMEs, they should 1) be familiar with the validation process and different validation techniques, 2) have at least a basic understanding of the different types of simulations and their purposes, and 3) be exposed to different types of data displays to help them prepare for the potential systems to which they could be exposed and help reduce misconceptions of simulation capabilities and intent.

Cognitive Task Analysis (CTA)

Cognitive Task Analysis (CTA) is an extensive/detailed look at tasks and subtasks performed by a person to achieve a goal. It seeks to describe the cognitive processes underling the performance of tasks and the cognitive skills required to respond appropriately to complex situations. [24] Thus, it examines actions and the decisions leading to those actions. Such an analysis could be used as bases for collecting the referent used for the development and validation of HBR. Because a CTA does not predict human behavior but outlines the human thought process, it can help to identify the factors individuals take into account when selecting a specific action. Such information could assist SMEs in the validation process by determining if the process used by a model is reasonable for the human behavior(s) the model is designed to replicate. Information from such a process could allow SMEs to determine if a model can be extrapolated for use in other situations in which referent is not available or for which the model was not evaluated. The information could also be used to identify open-ended requirements and limitations of human behavior specifications of domain specific situations and cultural bias.

The referent collected by such a process could potentially reduce the number of situations for which an SME would need to evaluate a model in order to gain significant confidence that the model was viable for its intended purpose, which purposes it could potentially be valid for, and which scenarios for which it would not be viable. As a result, time and money could be saved in obtaining the same level of validation currently achieved through the face validation of overt behaviors.

Human Performance Evaluation

Although HBR models are merely subsets of possible human performance considerations and actions, we can use human performance evaluation techniques during the validation process. Based on the model representation used and the level of validation one attempts to accomplish, HBR models processes and results could be categorized and evaluated based on one of three domains: psychomotor, cognitive, and affective. [25] Within these categories, there are levels of complexity that can be discovered and evaluated based on the types of actions and responses a model portrays.

Psychomotor addresses the model's ability to replicate physical capabilities. This could be analogous to the physical tasks the model can replicate and would be utilized in the evaluation of overt behaviors. If the model was designed to replicate human ground combat behaviors, reacting to indirect fire may be a skill one would expect the model to replicate. However, the ability for the model to replicate fighter pilot capabilities would be out of the normal spectrum of capabilities one would anticipate the model to be able to handle. Thus, efforts of SMEs could be focused on questions that look at the different levels of complexity with regard to these physical actions which could potentially reveal the extent of the model's capabilities to perform in the specified scenario or potential environments.

The cognitive domain refers to the model's algorithms that can give the validation agent an understanding of how the model determines which action to select. This category of evaluation could help the agent determine if the model could potentially perform correctly under scenarios not specifically tested. Evaluating this domain may not allow one to specifically say that a model will perform correctly, however it could potentially identify areas where a model will not be able to perform in a reasonable manner. This helps to identify areas for future testing and development when time and funds permit.

The affective domain is concerned with the emotional impact of individual values and priorities. Will an entity choose to perform a specific act if it has the mental and physical ability to do so? To date, most model representations have implemented theoretical models which have dealt with the physical and cognitive components of human behavior. As more theoretical human behavior models are implemented in code, the affective portion of validation will become more important.

Using CTA and human performance evaluation techniques would help model developers collect referent and validation agents develop questions to focus SME efforts. This focus could assist in correlating the evaluations of independent SMEs and potentially identifying areas of viable use of the model, while collecting relevant information for the development of future model modifications. Coupled with the classification, training and certification of SMEs these factors could improve consistency of results, reduce the number of SMEs required for each validation phase and allow for greater levels of validation for the same number of dollars.

CONCLUSIONS

Validation of cognitive models is a difficult process that is neither well defined nor uniformly complied with. The confusing and seemingly never-ending process of verifying, validating, and accrediting models for use in training and analysis of alternatives often leads the responsible agency towards a black hole of despair. Focusing on issues related to cognitive models to select areas of interest reduces the complexity to a more tractable problem.

Five issues with validating HBR models are incomplete or inaccurate referent, limitations of model representations, selection and use of SME's, limitations of face validation using overt behaviors, and the cost of the process. To address these issues, this paper suggests that using techniques from the fields of psychology and performance evaluation will improve the validation process for cognitive models. For models which allow the use of CTA techniques, a more extensive understanding of the information and processes used by the model to determine what actions to take can be extracted. This understanding will allow certain models to gain credibility for use in general situations. The use of performance evaluation techniques will help validation agents understand the different types of information and questions they can ask of a model and will focus their efforts and extend their validation to more general capabilities of the model. The classification, selection, training, and preparation of SMEs will help ensure competent and qualified validation agents are available to validate cognitive models. These factors will progress the VV&A process by producing more consistent results and expanding the level of understanding of HBR model capabilities.

FUTURE WORK

The principle recommendations proposed in this paper need to be further studied and implemented in the validation of a series of cognitive models to provide a proof of principle and credibility for their use. This work is currently underway at the MOVES Institute, Naval Postgraduate School, Monterey, CA. Preliminary results are expected by the Spring of 2004. [26]

REFERENCES

[1] Hollis, W. 2002. "M&S Expectations for T&E and the Challenge of VV&A". In *Proceedings of the MORS Mini-Symposium on Test & Evaluation, Modeling and Simulation and VV&A*(Kirkland AFB, Albuquerque, NM, 15 October 2002). Phalanx, The Bulletin of Military Operations Research, 1, 23-25.

[2] Pew, R. W., and A. S. Mavor, eds. 1998. *Modeling Human and Organizational Behavior: Application to Military Simulations*. National Academy Press. Washington, D.C.

[3] Department of Defense Directive (DoDD) 5000.59: Department of Defense Modeling and Simulation (M&S) Management. 1994. Alexandria, VA: Department of Defense.

[4] Office of the Director of Defense Research and Engineering (DDR&E), and Defense Modeling and Simulation Office (DMSO). 2001. Department of Defense Instruction (DoDI)5000.61 (Draft): Department of Defense Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A). Washington, DC: Under Secretary of Defense for Acquisition and Technology, Department of Defense.

[5] *Verification, Validation, and Accreditation (VV&A) Recommended Practices Guide (RPG): Glossary* [Website]. Department of Defense Modeling and Simulation Office (DMSO), 15 October 2001 [cited 24 January 2003. Available from http://www.msiac.dmso.mil/vva/Glossary/default.htm.

[6] Department of Defense (DoD) 5000.59-P: Modeling and Simulation (M&S) Master Plan. 1995. Alexandria, VA: Under Secretary of Defense for Acquisition and Technology, Department of Defense.

[7] *Verification, Validation, and Accreditation (VV&A) Recommended Practices Guide (RPG): Special Topic - Validation of Human Behavior Representations* [Website]. Department of Defense Modeling and Simulation Office (DMSO), 25 September 2001 [cited 24 January 2003. Available from http://www.msiac.dmso.mil/vva/Special_topics/hbr-Validation/default.htm.

[8] *Verification, Validation, and Accreditation (VV&A) Recommended Practices Guide (RPG): Special Topic - Validation* [Website]. Department of Defense Modeling and Simulation Office (DMSO), 15 August 2000 [cited 24 January 2003. Available from http://www.msiac.dmso.mil/vva/Special_topics/Validation/Validation.htm.

[9] *Verification, Validation, and Accreditation (VV&A) Recommended Practices Guide (RPG): Reference Document - Key Concepts of VV&A* [Website]. Department of Defense Modeling and Simulation Office (DMSO), 15 August 2001 [cited 24 January 2003. Available from http://www.msiac.dmso.mil/vva/Key/key.pr.pdf.

[10] Dean, T., J. Allen, and J. Aloimonos. 1995. *Artificial Intelligence; Theory and Practice*. The Benjamin/Cummings Publishing Company. Redwood City, California.

[11] Russell, S. J., and P. Norvig. 1995. *Artificial Intelligence: A Modern Approach*, *Prentice Hall series in artificial intelligence*. Prentice Hall. Englewood Cliffs, NJ.

[12] Schelling, T. C. 1978. *Micromotives and Macrobehavior*. 1st ed, *Fels lectures on public policy analysis*. W.W. Norton Company, Inc. New York, NY.

[13] Holland, J. H. 1995. *Hidden Order: How Adaptation Builds Complexity*. Edited by H. Mimnaugh. Perseus Publishing. Cambridge, MA.

[14] Ferber, J. 1999. *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence*. Addison-Wesley. Harlow, England.

[15] Roddy, K. A., and M. R. Dickson. 2000. "Modeling Human and Organizational Behavior Using a Relation-Centric Multi-Agent System Design Paradigm." Thesis thesis, Naval Postgraduate School, Monterey, CA.

[16] Lewis, T. G., M. J. Zyda, and J. E. Hiles. 2002. Proposal to Establish a Center for Study of Potential Outcomes. In *Modeling of Virtual Environments and Simulations (MOVES) Institute*. Monterey, CA: Naval Postgraduate School.

[17] Balci, O. 1997. "Verification, Validation and Accreditation of Simulation Models". In *Proceedings of the 29th Winter Simulation Conference*(Atlanta, GA, 07-10 December). IEEE, 135 - 141.

[18] *Verification, Validation, and Accreditation (VV&A) Recommended Practices Guide (RPG): Human Behavior Representation (HBR) Literature Review* [Website]. Department of Defense Modeling and Simulation Office (DMSO), 15 August 2001 [cited 24 January 2003. Available from http://www.msiac.dmso.mil/vva/Ref_Docs/HBR/beh-ref-pr.PDF.

[19] Matthews, G., D. R. Davies, S. J. Westerman, and R. B. Stammers. 2000. *Human Performance: Cognition, Stress*

and Individual Differences. Psychology Press. Hove, England.

[20] Pace, D. K., and J. Sheehan. 2002. "Subject Matter Expert (SME)/Peer Use in M&S V&V". In *Proceedings of the Foundations '02*(Lauarel, MD, 22-24 October). The Society for Modeling and Simulation International, 34.

[21] Vicente, K. J. 2000. Work Domain Analysis and Task Analysis: A Difference That Matters. In *Cognitive Task Analysis*, edited by J. M. Schraagen, S. F. Chipman and V. L. Shalin. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

[22] Pace, D. K., D. E. S. Stevenson, and S. M. Youngblood. 2002. "Foundations '02 Executive Summary". In *Proceedings of the Foundations '02*(Lauarel, MD, 22-24 October). The Society for Modeling and Simulation International.

[23] *Verification, Validation, and Accreditation of Simulation Models*. 2000. Huntsville, AL: AEgis Technologies Group, Inc. Course Matereials.

[24] Klein, G. A. 2000. Cognitive Task Analysis of Teams. In *Cognitive Task Analysis*, edited by J. M. Schraagen, S. F. Chipman and V. L. Shalin. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

[25] Hale, J. 2002. *Performance Based Evaluation: Tools and Techniques to Measure the Impact of Training*. Jossey-Bass. San Francisco, CA.

[26] Goerger, S. R. 2002. Validation and Evaluation of Cognitive Models for Combat Simulations. In *MOVES Institute*. Monterey, CA: Naval Postgraduate School.

AUTHOR

Simon R. Goerger is a Major in the United States Army and is currently assigned to the Naval Postgraduate School, conducting doctoral work in modeling and simulations. He has served the military as an infantry officer and successfully commanded light cavalry forces. He has been deployed and led soldiers in Egypt and Haiti as part of two separate peace keeping missions. He earned his masters in Computer Science from the Naval Postgraduate School prior to being assigned as a software engineer at the US Army's Training and Doctrine Command Analysis Center (TRADOC) at White Sands Missile Range, NM. During that assignment, MAJ Goerger worked on COMBAT^{XXI}, the US Army and Marine Corps next-generation stochastic, closed form brigade and below entity level combat simulation.