Proceedings of the 2003 Winter Simulation Conference
S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.

FIDELITY AND VALIDITY: ISSUES OF HUMAN BEHAVIORAL REPRESENTATION REQUIREMENTS DEVELOPMENT

Tom Hughes
Evan Rolek

General Dynamics, Advanced Information Systems
5200 Springfield Pike, Suite 200
Dayton, OH 45431, U.S.A.

ABSTRACT

Within the modeling and simulation community issues of fidelity and validity are often considered the fundamental metrics used to gauge the quality and utility of a model or simulation. The extent to which a model represents reality is often referred to as fidelity while the usefulness of that representation within the context of a specific application relates to its validity. In most cases, increases in representational fidelity are coincident with increased development costs. How then does one decide where investments in fidelity should be placed? Faced with this question regarding the representation of operator behavior for a synthetic integrated air defense system (IADS) to be used in a synthetic battlespace, we are conducting a requirements analysis to establish human behavioral representation (HBR) requirements. This analysis will provide a degree of traceability from fidelity requirements to simulation objectives so effective tradeoffs can be made. The present paper reviews our methodology and preliminary findings to date.

1 INTRODUCTION

Modeling and Simulation (M&S) has become fundamental to military training, analysis and systems acquisition. The utility of a M&S approach is based on perceived efficiencies over traditional methods, particularly in the areas of training and system acquisition. A large-scale live training exercise can cost the Department of Defense (DoD) several millions of dollars in manpower, equipment and expendable resources, not to mention the safety risk that is inherent in every live exercise. The same training effect can reportedly be achieved through distributed simulation techniques for a fraction of the costs and without the associated risk. Likewise, the business of weapon system acquisition has been radically transformed through the systematic application of M&S tools and techniques. Developmental questions that remained unanswered until a test vehicle was made available can now be investigated within a synthetic battlefield environment. Technology and cost tradeoffs can be studied in great detail well before large investments in system fabrication are made. These efficiencies are driven by the increasing trends that require the military to continue to do more with less. As we come to depend more and more on the capabilities and insights afforded by M&S tools it is time that we critically assess the quality of these tools and the processes by which they are created and used. Only through such a process can we ensure that our faith in this approach is well founded.

A fundamental assumption upon which the use of M&S is being advocated is that current techniques provide an adequate and cost effective surrogate for reality. However, it is an inescapable truth within the M&S community that "all models are wrong, but some models are useful." (AGARD Aerospace Medical Panel, 1998). All models are wrong because a model is not reality, but merely a limited, (and in this context *computational*) representation of reality. As with any model, the quality of the model depends upon how well those that developed the model understand the reality it supposes to represent. With regard to physical systems, computational models can provide a level of precision that is remarkably exacting within the boundaries of their scope and underlying assumptions. Unfortunately, this level of precision does not transfer to representations of the human component of these systems. Traditionally, representations of human operators have been relatively ineffectual as a result of oversimplified assumptions underlying the models. It is generally agreed that the most critical -- and most complex -- component of any weapon system is the operator. Ironically, the limited degree to which crew behavior is accurately represented in these simulations is generally regarded as inadequate, and as such, limits the overall validity and utility of the models.

A major review of the current state of human behavioral representation notes that, in general, "user expectations exceed Human Behavioral Representation (HBR) capabilities" and "HBR falls short of user needs." (Pew and Mavor, 1998) In light of these facts, the Defense Modeling and Simulation

Office (DMSO) has identified an increased robustness in the representation of individual and group behaviors as a critical need. DMSO objectives specifically state the need to (1) extend existing models of combat operations to include individual combatants, (2) develop generic models of individual human capabilities, limitations, and performance (physiological and psychological), and (3) develop the capability to rapidly construct models of individual human behavior for specific applications on demand.

Often what is at issue is what it means to be robust and more importantly, what makes a particular representation valid. How does one define fidelity within a simulation environment? How does one establish fidelity requirements for a given simulation exercise, and how is that level of fidelity implemented in a representation of human operator behavior? This paper attempts to address several of these issues as they relate to the development of human performance representations within military modeling and simulation. Our specific application deals with modeling to support acquisition decision-making processes. We first discuss issues of fidelity and validity by reviewing some of the findings from the Simulation Interoperability Workshop's Fidelity Integration Study Group (ISG). We then propose a framework for the development of Functional Description of the Mission Space (FDMS) focusing on human operator performance within a specific context. We provide an example of this framework for the development of a human operator FDMS for a ground-based air defense system. Finally we describe an approach for applying the FDMS to support the development of simulation requirements based on the specific research questions and issues of interest.

2 FIDELITY AND VALIDITY

2.1 Definition

Model fidelity and the level of detail captured in a model's representation of entities continue to be major issues in the development of simulation models. The Fidelity ISG of the Simulation Interoperability Workshop conducted an extensive review of fidelity issues associated with military simulations. Their review has provided a working definition of fidelity that we will use in our discussion. The working group defines fidelity as "The degree to which a model or simulation reproduces the state and behavior of a real world object or the perception of a real world object, feature, condition, or chosen standard in a measurable or perceivable manner; a measure of the realism of a model or simulation; faithfulness. Fidelity should generally be described with respect to the measures, standards or perceptions used in assessing or stating it." (Gross et al, 1999). Simply stated, fidelity can be described in terms of the extent to which a representation reproduces the attributes and behaviors of a referent. In this case a referent is an entity or collection of entities and/or conditions -- together with their associated attributes and behaviors -- present within a given operational domain.

While it can be argued that fidelity can be quantitatively described, validity, in our opinion cannot. As we have just discussed, fidelity is an objective assessment of a representation's capture of the attributes and behaviors of a referent. Validity, on the other hand, is a judgment regarding how well suited a particular representation is for a *specific* application. The word *specific* is highlighted because validation is a judgment that must be conducted for every application of a given model or representation. A model valid for one application is not necessarily valid for another. The distinction lies in the particular behavioral phenomenon that is being investigated within the simulation. A relatively low fidelity representation of an entity's attributes and behaviors may prove valid if its behavioral phenomenon is not critical to the system level phenomenon of interest. Therefore, while fidelity and validity are related, validity does not necessarily imply fidelity and fidelity does not guarantee validity. Fidelity can be measured, validity must be judged. It is for this reason that those charged with assessing the validity of a representation for use within a given simulation must possess an intimate understanding of the physical reality of the entity being represented as well as the goals and objectives of the simulation exercise. Investments in increased fidelity when lower levels of fidelity could provide a valid simulation result in unnecessary expenditure of valuable resources. Therefore, the representation requirements analysis process is essential to a cost efficient simulation development effort.

2.2 Referent

In military modeling and simulation, those agents and entities (together with their attributes and behaviors) present in the operational environment ideally define a referent. To facilitate the transfer of domain specific information from the operational domain to the development of computational models, analysts first develop a repository of authoritative knowledge about what is known about the mission space. Our review of a select number of FDMSs revealed that these descriptions tend to primarily focus on systems, and emphasize the physics of system operations. Like many of the models we investigated, they provide an extremely limited treatment of human operator functionality and capability. Given that the FDMS is designed as a knowledge repository to inform modelers, it becomes evident why many models provide such a limited treatment of human behavior; their referent lacks the necessary knowledge to inform the modeler.

As developers and users of models, we must remain sensitive to the relative costs associated with various levels of model fidelity. As one strives for increases in simulation fidelity, costs associated with these increases typically escalate. It is for this reason that study managers are interested in answering questions regarding how much fidelity is necessary to effectively respond to established research issues. Figure 1 presents a notional relationship between simulation fidelity and simulation investment. Conceptually, this relationship can be represented as an asymptotic function. Initial increases in investment will result in substantial increases in fidelity. At some point, continued investments will result in diminishing returns in fidelity.

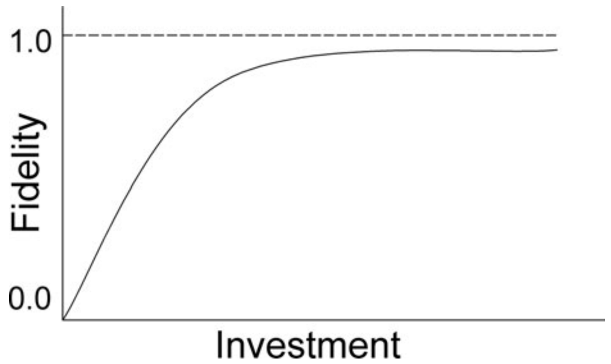

Figure 1: Fidelity is the Degree to Which a Model Reproduces a Referent as Modified from Gross et al. 1999

2.3 Measurement

Some may question whether fidelity can actually be described in quantitative as well as qualitative terms. The Fidelity ISG has argued that quantification of fidelity in not only possible -- but also necessary -- to achieve effective interoperability between simulations. Traditionally, fidelity has been described in extremely vague terms (e.g., high, medium, low) with no real common understanding of what such designations meant. With regard to human representations, the challenge is first to describe the referent (i.e., the human operator) and the dimensions along which fidelity can and should be described. This task, in itself, is a daunting one. Once this is done, fidelity can be defined in terms of the attributes, behaviors and processes that a given representation is able to reproduce relative to an established referent.

3 DEFINING FIDELITY REQUIREMENTS

As is often the case in other development programs investments in requirements development are often limited. Resource constraints and delivery deadlines often foster a sense of urgency to initiate implementation of a model. Dedication to a structured requirements development process can have significant payoff in two specific areas. First, given a thorough understanding of the study's objectives and associated modeling requirements, investments in unnecessary fidelity can be avoided. Chandrasekaran and Josephson (1999) presented a convincing argument for the need to establish a structured requirements process to support the development of valid human representations in military simulations. They argue that "the pursuit of high fidelity cognitive models, unfettered by detailed considerations of what we want the models for, is so unfocused as to be almost useless for practical purposes." They have proposed a requirements definition process for establishing human representation fidelity requirements. This approach focuses on the conduct of detailed analysis of operator functions and tasks as a means of gaining insight into crew operations within the operational domain. Equipped with these insights, analysts are in a much better position to make sound judgments regarding the level of fidelity required to respond to training, analysis or systems acquisition issues established by simulation objectives.

The second area where a structured requirements process can pay significant dividends is determining where increases in model fidelity will be necessary. Insights that inform analysts where low levels of model fidelity may be sufficient can also guide analysts in assessing the need to enhance representational fidelity. As the DMSO Master Plan Objective 4 suggests, representations of individual combatant behavior are lacking. The question analysts must answer is in what area are they lacking. Intuition suggests that all models are not lacking in the same way. Necessity dictates that different simulation efforts will have significantly different fidelity needs. Identifying when and how these needs differ is the challenge faced by both model developers and model users. The application of a structured model requirements approach can guide investments in model development that will focus on those aspects of human behavior and performance that are most sensitive to the alternative conditions that are to be studied.

A process for fidelity requirements development is presented in Figure 2. As the figure suggests, operational domains are defined by the physical reality in which they exist.

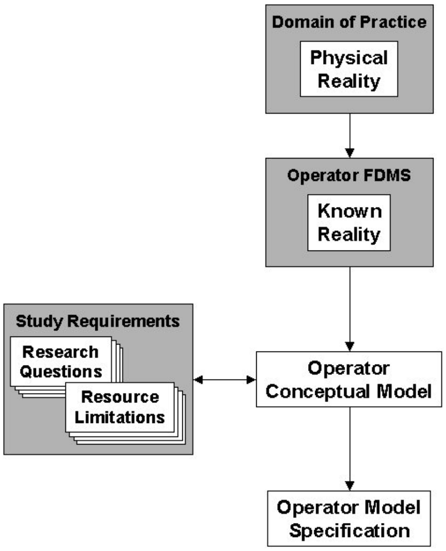

Figure 2: Operator Model Fidelity Requirements Process

Unfortunately we oftentimes never know all aspects of the physical reality and can only describe the known reality reflected in the real world. As a result, there will always be levels of detail about real-world systems that will remain unknowable. The question that must be asked is how much insight is necessary? It is this known reality that we can attempt to capture in the FDMS. Recognizing that the depth and breadth of the known reality the analyst must develop a taxonomy or catalogue to organize and scope the capture of domain information and knowledge. The proposed Operator FDMS framework can provide the necessary structure for data capture. Once initiated, the Operator FDMS can serve as a repository for the body of knowledge regarding operator behavior and performance within the domain of practice. This body of knowledge can be made available to other model developers to serve as a point of departure of the development of other conceptual models. Over time the FDMS can continue to grow in breadth and depth and serve as an authoritative resource of information.

By mapping insights gained through the development of the FDMS against research questions and resource constraints defined by the study requirements we can begin to develop an operator conceptual model that will be used to define specific operator model requirements. We may discover that although a particular attribute of crew performance has high priority as defined in the FDMS, it may be completely insensitive to the range of conditions defined by the study objectives. Under such cases it may be acceptable to provide lower levels of fidelity for this attribute, even though it may be considered critical, and focus resources in developing fidelity in areas that are sensitive to the research conditions. The analyst must perform these trades to ensure that available study resources provide for the most value added to the study effort. The results of these trade analysis will be an operator model specification that will direct the implementation and application of the human operator representation to be used in the simulation exercise.

We have adopted an approach and methodology that is designed to augment current and future FDMSs with detailed analyses and descriptions of the operator characteristics associated with systems operations. We envision a three dimensional framework for providing such a description. Depicted in Figure 3, this framework describes operator functionality in terms of (1) a operator mission functions, (2) operational constraints, and (3) human behavioral processes. For each cell within the framework, the analyst -- utilizing an analysis of the mission space, SME input, direct observation, and data collection -- provides a detailed description of the human operator's contribution to system operations and performance. Obviously, the design of operator interaction varies significantly across multiple mission tasks. Likewise, human behavioral processes play different roles based on the nature of the mission task and the operational constraints imposed by technology, tactics, and operator skill and experience. Therefore, the relative importance of any given cell defined within the framework can vary significantly.

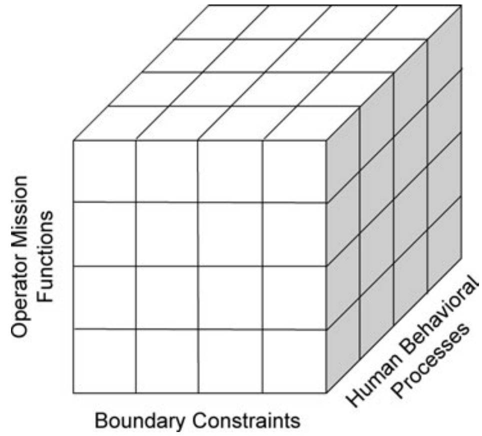

Figure 3: Human Operator FDMS Framework

4 A CASE STUDY: GROUND-BASED AIR DEFENSE ENGAGEMENT MODEL

Developing fidelity requirements for the representation of a surface to air missile (SAM) crew is a multi-dimensional process that embodies the combined expertise of operations analysis, subject matter experts from the operational community, and the human factors discipline. Figure 4 illustrates some of the dimensionality associated with describing the mission space from a crew centric perspective. Simplifying assumptions have been made and some details have been left out in the interest of brevity. For example, crew activities are not performed independent of one other. While their presentation may suggest serial execution, mission function can occur serially, in parallel and iteratively. Further, a "failure" for one activity, such as losing the target during tracking, can trigger a reversion to a previous activity, such as target search. Similarly, human behavioral processes such as sensation, cognition, motor functions, etc., interact and are interdependent with one another. The challenge of systems analysts is to understand and capture the nature of these actions and interactions relative to the mission functions and activities as they occur within the context of operational engagements.

During the process of prioritizing fidelity requirements, there are a number of factors that should be taken into account including the rationale for the crew representation, questions being addressed, assumptions made, level of resolution/detail for the crew representation, etc. The framework depicted in Figure 4 is developed based upon a decomposition and analysis of crew functions associated with systems involved in a given engagement or mission simulation, in our example, and might include the SAM crew functions as well as those performed by the air crew being engaged. The factors just mentioned exert an influence when filling in the table entries (i.e. low, medium, or

		Analyst's Visualization of SAM Crew?					
		comm with C2	ROE (bounds tactics and decision making)	tactics	decision making (incl. supervisory control)	skill	crew interactions
SAM Crew Activities	activation	H	L	L	L	L	L
	search	H	M	L	H	H	M
	detect	M	H	L	H	H	M
	acquire	L	M	H	M	H	H
	recognize	L	M/H	L	L/M	H	H
	identify	L	M/H	L	L/M	H	H
	prioritize	L/M	H	L/M	M/H	L/M/H	L/M/H
	track	L/M	H	H	H	H	H
	launch	L/M	H	H	H	M	H
	flyout	L	L	H	H	H	H
	endgame BDA	H	L	L	M	M	M

sensation
perception
cognition
motor
comm

Figure 4: Conceptual Model Framework for Ground-Based Air Defense System Operator

high) criticality for that particular element of crew involvement. The completed table would then depict a mapping of SAM crew activities versus an analyst's visualization of the crew, along the dimensions of sensation, perception, motor functions, etc. A composite scoring method could then be implemented to identify the relative importance for each of the cells in the table thus reflecting the most important aspects of fidelity for the crew representation. A few cells in the table will be discussed for illustrative purposes. Note that this is a sample case where we're dealing with a command guided SAM, i.e. the radar must continually track the target and send guidance commands to the missile for aiming it at the target aircraft. Interactions between "Track" crew functions and "Skill" was rated as "H" for high overall crew involvement. While radar systems contain autotrack loops negating the need for manual tracking by the crew, it is often the case that electronic countermeasures force the crews into manual tracking modes. Unless this tracking function is accurately represented, quantifying the performance of the overall SAM system will be in error. This is particularly true for a command guided missile system since any tracking error will ultimately manifest itself as an error in the estimated target position being sent to the missile for intercept. The skill level of the crew will determine how much error there is with estimating the aircraft's location.

Interactions between the "Launch" function and "Decision-Making" was also rated as "high" overall. This particular set of activities is one of the most critical for a SAM crew during engagement of an aircraft. Decisions include determining the optimum launch mode, missile guidance mode for each missile, warhead arming, etc. prior to the launch of each respective missile. Poor decisions can result in the aircraft exiting the engagement zone prior to intercept, allowing the aircraft to come too close to the SAM for engagement, losing track of the aircraft, or falling victim to an attack against the SAM system itself by the aircraft.

Similarly, interaction between "Flyout" functions and "Decision-Making" was rated as "high" overall because of the criticality of crew activities here. During the missile flyout, the crew must constantly optimize the selection and implementation of tracking modes used to track the target in azimuth, elevation, and range with respect to the SAM site. In this age of information warfare, electronic countermeasures are commonplace. The crew must use available electronic counter-countermeasures (ECCM) to negate the effects of the jamming during the time critical activities associated with missile flyout which, for some systems, can have a duration of less than 20 seconds from launch to intercept. Hence, decisions must be made quickly and accurately to ensure a successful intercept. In addition, the crew must constantly assess the situation and make decisions about how many missiles to launch against the penetrating aircraft.

After all of the cells in the table have been assessed relative to their "criticality" analysts can prioritize those areas upon which to focus knowledge elicitation and data collection efforts. In some cases, nominal values may be pursued that are indicative of "average" crew performance. However, alternative types of crew performance include expert performance, worst case performance, performance indicative of novice through seasoned operators, irrational or unexpected acts, etc. and it is up to the study team to specify what characterization is most important for the respective study. In some cases, the team might decide to use an available performance data base that was obtained from a laboratory simulation, field test, or combat experiences. While some of these decisions are indeed difficult, they can be approached in a systematic manner by using the FDMS process advocated here.

We used the insights into the air defense domain gained by this framework to assess the representation of system operator representation within an existing air defense engagement model, the Enhanced Surface to Air Simulation (ESAMS, 1997). The ESAMS series of models are traditionally used as one-on-one engagement level models. That is, a single penetrator is flown against a single air defense site. Although not intended as a one versus many or many versus many engagement model, the model can be used for such applications through numerous setups and execution of the model. Using an engagement model to approximate a mission level model is sometimes done in very specific applications but is, in general, not recommended due to the inefficiencies of such an exercise and the fact that the analyst must manually adjust the engagement model to take into account mission level factors. ESAMS currently models many different surface-to-air defense systems. In addition to selecting the particular threat system modeled, a broad range of other engagement factors can also be manipulated. Crew capabilities, however, are not overtly considered for either the penetrating aircrew or the crew of the air defense system.

Since ESAMS is often used to assess the impact on survivability, we describe two areas where the lack of appropriate insight into crew operators may prove problematic in providing accurate estimates of survivability rates. "Tracking" and "ECCM" are two functions where crew participation in the air defense system process is critical.

During target tracking, there is an error that represents the accuracy of the tracking system that, in the case of the crew, is a function of the operator(s) skill and the characteristics of the servo systems used to position the radar antenna. Although not technically pure, it is possible to take the predefined flightpath of the aircraft and add to it a sinusoidal (or sum of three sine waves) term for altitude and lateral position. This can be used as an approximation of the results of a SAM crew tracking the aircraft when in fact a radar autotrack mode is used within the model. Ideally, it would be more precise to have a human operator model of tracking performance; such models have been developed and validated for some systems.

When performing ECCM activities, the crew will select various system capabilities to counter the jamming that is present; these capabilities can include changes to the radars frequency, pulse rate, pulse waveform, etc. With the programmable logic available within ESAMS, it is possible to replicate the time it takes the crew to activate various ECCM capabilities and essentially "turn off" the jamming and/or reduce the amount of tracking error caused by the jamming.

5 SUMMARY

While we attempt to make advances in providing robust representation of human behavior and decision-making performance, it is essential that we remain grounded in support of established research objectives. Pursuit of high fidelity representations of human behavior that do not directly support the research objective may divert resources away from areas where real impact can be realized. Understanding where and how increases in fidelity can have an impact on the validity of the model can only occur as a result of a thorough and accurate assessment of the operational domain. We have outlined a framework for the capture of such insights as it relates to human operator involvement in system operations. Our future efforts will focus on the application of this process within a specific modeling effort to expand on the knowledge repository and use the information to instantiate a valid representation of human performance of appropriate fidelity.

REFERENCES

AGARD Aerospace Medical Panel 1998. A Designer's Guide to Human Performance Modeling. AGARD-AR-356, NATO/AGARD, Neuilly-sur-Seine Cedex, France.

Chandrasekaran, B. and Josephson, J. R. 1999. "Cognitive Modeling For Simulation Goals: A Research Strategy for Computer-Generated Forces" 8th Computer Generated Forces Conference, Paper No. 8TH-CGF-100.

ESAMS Analyst's Manual, BDM/ABQ-97-0023-TR, August 15, 1997, BDM International, Inc., Contract F41621-96-D-5001, Delivery Orders 12 and 13, CDRL A00D, unclassified.

Gross, D., Pace, D., Harmon, S., and Tucker, W. 1999. "Why Fidelity?" Spring 1999 Simulation Interoperability Workshop, Paper No. 99S-SIW-168.

Pew, Richard W. and Mavor, Anne S. (Eds.) 1998. Modeling Human and Organizational Behavior: Application to Military Simulations. National Research Council. National Academy Press, Washington DC

AUTHOR BIOGRAPHIES

TOM HUGHES is a Program Manager with Veridian Engineering in Dayton, OH. He received his BS in Industrial and System Engineering from the Ohio State University, a M.S. in Psychology from Wright State University, and is currently pursuing a Ph.D.. .in Cognitive Systems Engineering from The Ohio State University. He has 21 years experience in research and development with emphasis in the area of human systems integration and human performance modeling. Prior to joining Veridian Mr. Hughes spent 12 years with the Air Force Aeronautical Systems Center where he served as Program Manager and Principle

Investigator at the Crew Station Evaluation Facility, a pilot-in-the-loop simulation facility that specializes in the evaluation of advanced crew system concepts and the investigation of crew integration related issues. His email address is <tom.hughes@wpafb.af.mil>.

EVAN ROLEK is a Senior Scientist with Veridian Engineering in Dayton, OH. He received his B.S. in Engineering Psychology from the University of Illinois. He has 35 years of combined experience in the areas of human performance assessment and modeling, simulator development, and field testing of military and commercial systems. Areas of experience include command and control, air defense, air crew performance, electronic warfare, and control and display evaluation for military and automotive applications. Prior to his current affiliation, he was with Science Applications Incorporated Corporation, Systems Research Laboratories Incorporated, and General Dynamics Electronics Division. His email address is <evan.rolek@veridian.com>.