# VALIDATION OF HUMAN BEHAVIOR REPRESENTATIONS

## RPG Special Topic

## 9/25/01[1]

### Table of Contents



[1] This document corresponds to the 8/15/01 online version of the same name. It contains only minor formatting changes.

*This document corresponds to the web version of the special topic of the same name and date. It has been modified to make it suitable for printing.*

## Table of Contents Continued

## Table of Contents Continued

# Introduction

This special topic paper discusses the considerations unique to validating human behavior representations (HBRs) and the tools and techniques developed to support these activities. This discussion is divided into seven topical areas, which can be accessed via hot link:

- **Basic HBR validation concepts** – a variety of issues that introduce the reader to HBR validation
- **HBR requirements** – discussion of the issues associated with acquiring a set of HBR requirements sufficient for validation
- **Referents for HBR validation** – guidance for identifying the referents needed to support HBR validation
- **Subject matter experts** – several issues guiding the reader in the use and management of SMEs specifically for HBR validation
- **HBR conceptual model validation** – guidance for validating HBR conceptual models
- **HBR knowledge base validation** – guidance for validating HBR knowledge bases
- **HBR results validation** – guidance for developing and executing the results validation process for simulations involving one or more HBRs

The discussion in each topical area is formatted as a series of questions, which are listed at the beginning of the area and can be accessed via hot link. This format allows the reader to select those sections that seem pertinent to the prevailing problems without having to digest the entire document from beginning to end. The discussion associated with each question contains sufficient information to answer that question. Since questions often address overlapping areas, this organization has lead to some necessary redundancy. This redundancy may annoy a reader consuming this document serially but it significantly reduces the burden on the reader accessing this information randomly, the mode in which the authors anticipate the most use.

# Basic HBR Validation Concepts

HBRs occupy a unique niche in the modeling and simulation (M&S) arena. Thus, they bring a distinct set of problems to validation. This section provides a brief and broad overview of HBRs and how validation fits within their niche. It addresses the following questions:

- What is a human behavior representation?
- How do HBRs work?
- What are behavior moderators?
- What distinguishes HBRs from other simulation types?
- What makes validating HBRs difficult?
- What does HBR validation generally involve?
- What is the most common form of HBR validation and why?
- Why doesn't face validation adequately answer the validation questions?
- What development paradigm best suits HBR validation?
- How do the learning capabilities of an HBR affect its validation?
- What tasks must one perform to validate an HBR?
- What information does an HBR validation plan contain?

These questions serve as the springboard into the remaining sections and they, where appropriate, guide the reader to the more detailed information in these later sections.[2]

## *What is a human behavior representation?*

All HBRs model the behavior of people at some level. The term HBR encompasses representations of parts of individuals (e.g., hands operating controls), individuals (e.g., a specific terrorist or equipment operator), aggregates of individuals (e.g., a crowd, a command staff), and aggregates of organizations (e.g., several organizations responding in concert to an emergency situation).

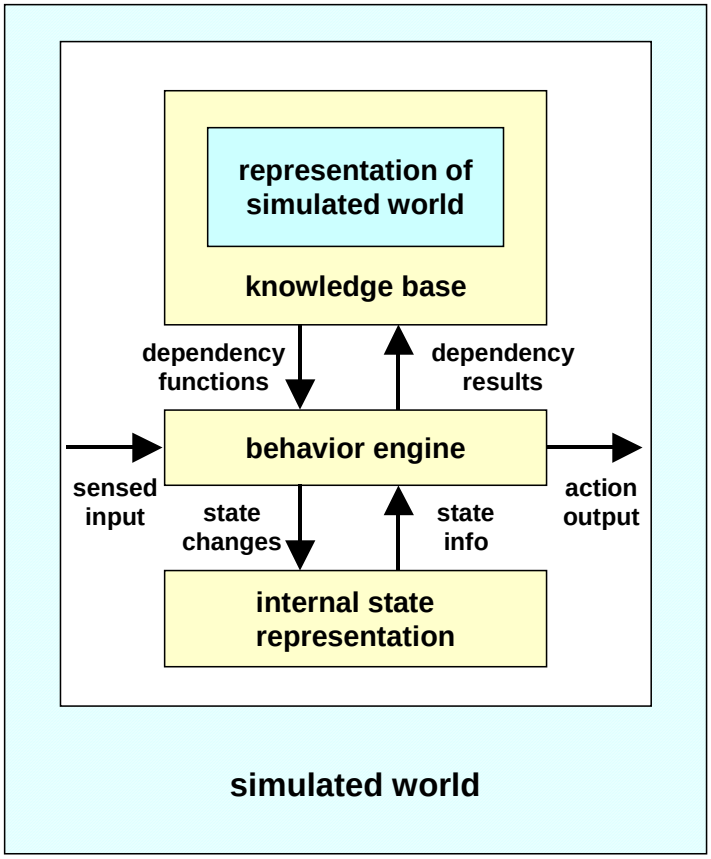
An HBR may depict one or more classical cognitive functions (e.g., perception, inference, planning, control), human performance limitations (e.g., sensing bandwidth, decision latencies) and the effects of behavior moderators (e.g., stress, injury, fatigue, discomfort, motivation and emotion). HBR implementations vary from simple finite state machines to complex knowledge-based systems integrating multiple reasoning paradigms and augmented by simulations of the effects of various behavior moderators. The terms computer generated forces (CGF), semi-automated forces (SAF and SAFOR), synthetic forces, automated forces (AFOR) and command forces (CFOR) all refer to different forms of HBRs.

## *How do HBRs work?*

The figure below illustrates the components of a generic HBR. This illustration, derived from a more comprehensive model [Harmon, 1999], depicts the classical knowledge-based system (KBS) architecture consisting of an execution engine interacting with a

[2] See the reference document on HBR Literature Review for additional information.

knowledge base to update an internal state representation. Since the execution engine of an HBR produces its observable behavior, it is called the ***behavior engine***. Most HBRs follow this general architectural path, at least conceptually, in order to partition the contents of their knowledge bases from the mechanisms needed to execute them. This partitioning creates the flexibility needed to represent the behavior of different individuals performing in different roles without requiring building a new execution infrastructure each time. This partitioning can still be applied to HBR designs that do not explicitly adopt it. For example, human neural anatomy and physiology appear to tightly integrate knowledge within the execution mechanism yet that physiology varies relatively little from individual to individual even though the knowledge they possess and the behavior they manifest may differ enormously.




**Description of a Generic Human Behavior Representation**

5/23/01

Models of sensors and effectors (not shown in the figure) couple the HBR to the simulated world. ***Sensors***, representing things like eyes, ears and nose, receive information about the world and ***effectors***, representing things like hands, arms, legs and voice, change the state of the world in some observable way. Sensor models deliver sensed information to the behavior engine.

The behavior engine, at a minimum, performs the following functions:

- accepts input about the state of the simulated world and uses that information to update its own internal state representation
- chooses additional relevant information from its knowledge base to augment its internal representation of world state
- applies its current assessment of world state to choose and execute information from its knowledge base to generate the responses it thinks can lead to achieving its goals

The internal state representation contains the perceptions, goals, and plans relevant to the current situation. The behavior engine generally represents the HBR's reasoning abilities independent of the particulars of a situation. Behavior engines may also represent the influences of various behavior moderators and performance limitations (e.g., processing bandwidths) if the knowledge base does not completely encode them.

In learning systems, the behavior engine also compares its internal state representation against the information contained in its knowledge base to derive modifications to the knowledge base itself. The changes to the internal state representation and the knowledge base together with the output it produces completely represent the HBR's path through its behavior space. In other words, in the context of this model, an HBR can only change the contents of its internal state representation, knowledge base, and action output to manifest its behavior.

In essence, the behavior engine determines these changes by choosing the relevant knowledge from the knowledge base and applying that knowledge to the current internal state representation. The particular mechanisms it uses to perform these operations depend upon the knowledge representation scheme chosen. The knowledge base represents all of the persistent information that the HBR possesses. It encodes elements of the HBR's knowledge in the terms of one or more knowledge representations (e.g., production rules, semantic networks, task frames, probabilistic networks, neural networks). The complete content of the HBR's knowledge base, in effect, defines the computer program that, through the behavior engine, generates the behavior of the individual or group represented.

## *What are behavior moderators?*

The term behavior moderator refers to those phenomena that affect human behavior other than the cognitive elements. Pew and Mavor [1998] broadly classify behavior moderators into external and internal moderators. External moderators include physiological stressors and cognitive workload stressors. Physiological stressors refer to environmental factors (e.g., temperature, toxic substances, noise, vibration), physical workload, and fatigue. Cognitive workload stressors represent the overloading of various sensor input and computational bandwidths of the human. Internal moderators

are intelligence, expertise, cognitive abilities, personality, emotions, attitudes and cultural values. All of these factors, both internal and external moderators, change the base performance of a human's behavior in various ways; ways that probably vary from individual to individual.

## *What distinguishes HBRs from other simulation types?*

At first blush, the differences between HBRs and other types of simulations come primarily from their very high inherent complexity. Human behavior arises from numerous nonlinear relationships all interacting chaotically over many different orders of magnitude. Complex coupling with other parts of a simulation system may also seem to differentiate HBRs from other simulation components. However, simulated environment and nuclear effects models both face similar problems.

The real distinction of HBRs comes from their knowledge bases. Almost all, if not all, HBRs incorporate some form of knowledge base and an engine that executes the elements of that knowledge to produce the observable behavior. This arrangement permits using the same execution engine to represent different personalities functioning in many different situations. This reuse convenience means that the knowledge base really constitutes a computer program, in many cases a very complex one, which the engine executes.

An HBR's knowledge representation defines its programming language. A simulation system that represents different humans contains many of these computer programs within computer programs. Further, this distinction means that developers must diagnose and debug two sets of computer programs, the engine (usually written in a conventional programming language such as C) and the knowledge base for each individual represented (written in the knowledge representation language). These facts added to the inherent complexity of HBRs easily make them the most complex components of a simulation system, even when compared to simulated environments.

## *What makes validating HBRs difficult?*

Advances in information system technology make the construction of fine-grained simulations of human behavior for a variety of situations more feasible and practical. Developers have built many primarily cognitive simulations of human behavior and the sophistication and fidelity of these systems continues to improve. However, the technology for validating HBRs has not kept pace with these advances despite facing particularly vexing problems since their very first applications. Human behavior manifests an intricate fabric of effects coupled over many orders of magnitude, a property shared by complex chaotic systems. Small situation changes often create wildly different responses in the same system. Thus, validation of HBRs, even for simple tasks, can prove extremely difficult because of the large number of behavioral paths that must be explored for any given purpose. The lack of well-established techniques and tools to support HBR validation further exacerbates the difficulty of these problems.

## What does HBR validation generally involve?

At the most abstract level, validating an HBR involves comparing that HBR's capabilities to the M&S requirements for HBR in order to determine its fitness for the purpose represented by those requirements. Usually, defining HBR capabilities involves testing the HBR within the simulated environment in which it will be used and determining if the behaviors manifested in test results are sufficient. This testing should examine the functionality of both the HBR's behavior engine and of its knowledge base, differentiating the problems in each as much as possible. However, this minimalist view of HBR validation, while comprising an essential step, oversimplifies the validation process and, if executed alone, will likely result in either an incomplete validation effort, a costly validation effort, an effort that exceeds its schedule, or some combination of these three outcomes. Development of a good HBR validation plan can help to manage the validation process and avoid many of these difficulties.

At the most general level, to validate an HBR one should

- develop an adequate statement of requirements
- identify the referents that define the standards for determining accuracy or error
- assess the capabilities of the HBR
- compare those capabilities against the requirements to determine the fitness of the HBR for the purpose

One can assess the capabilities of an HBR at different phases of its development including

- design, as described by the conceptual model
- knowledge base contents
- engine and knowledge base implementation
- integration with its simulation environment

However, the accessibility of these different phases depends strongly upon the knowledge representation employed. Some knowledge representations, such as neural networks and genetic algorithms, currently only permit their evaluation at the conceptual model and integrated levels as black boxes.

## What is the most common form of HBR validation and why?

By far, face validation is the technique most often applied to HBRs.[3] In this technique, a subject matter expert (SME)[4] drives the HBR through the scenario space by issuing

[3] See the reference paper on V&V Techniques for additional information.

commands or changing the stimulating situation, observes the resulting behavior, and determines, often qualitatively, whether that behavior meets a User's requirements for realism.[5]

## *Why doesn't face validation adequately answer the validation question?*

Regrettably, face validation is the least reliable and least complete HBR validation technique. HBRs can easily create extremely large and convoluted behavior spaces. Unlike many physical models, HBR sensor, effector and behavior generation mechanisms have very large sets of interdependent variables. Even simple HBRs can differentiate thousands of situation conditions and produce hundreds of responses to those conditions.

Highly nonlinear relationships between perceived situations and derived responses to those situations are commonplace. This nonlinearity means that one cannot generalize the behavior observed (and perhaps validated) for one set of conditions to another set even though the differences between the conditions may be small and insignificant. Further, unlike most environmental models, HBR knowledge bases represent executable information and may contain the same amount of information as moderately complex terrain representations. Yet, few would consider face validation of complicated physical or environmental system simulations adequate even though they are computationally simpler than most HBRs for realistic purposes.

## *What development paradigm best suits HBR validation?*

The complexity and uncertainty associated with any HBR and its validation makes any iterative development paradigm (e.g., rapid prototyping, spiral) appropriate.[6] These paradigms all encourage iterative formulation of requirements, acceptability criteria, the conceptual model and, ultimately, the HBR itself. They provide the most flexibility in dealing with understated requirements and in managing development risk. The evolution of these artifacts throughout the development effort demands iterative validation. While this approach to validation is not as immediately satisfying, it permits all of the involved roles to refine their contributions in response to feedback from the other roles. This is particularly important between the V&V agent and the developer. An iterative approach to HBR validation maximizes the chance for anticipating and planning a complete and consistent set of behaviors for the purpose.

The mention of development paradigm may suggest that this advice applies primarily to new development efforts. However, an iterative approach to using a legacy HBR is equally important. Users seldom sufficiently state their requirements for HBRs and these requirements constantly evolve as the Users exploit the power of the HBR. In

[4] See the special topic on Subject Matter Experts and VV&A for additional information.
[5] See the special topic on Validation for additional information.
[6] See the special topic on Paradigms for M&S Development for additional information.

these cases, showing the Users the HBR's capabilities can impact their requirements. Requirements changes should provoke further validation activity and so forth until the Users get what they really need.

## *How do the learning capabilities of an HBR affect its validation?*

Learning systems present special challenges since their observable behavior changes with each successive exposure to different world situations. As a result, validating the behavior of a learning system at different points in the development of its experience base could produce different results. Learning system behavior can be validated at two levels:

- Examining the correspondence of the HBR learning algorithms against the current theories of learning mechanisms and their properties
- Validating the HBR's behavior after its performance has stabilized

The first level assumes that the learning algorithms have been derived from some theory of human learning. The second level assumes that the system has been exposed to enough operational scenarios that its knowledge base has reached some point of stability. Unfortunately, if during its applications, a learning system is exposed to new conditions then its validity may stray, perhaps significantly.

Results validation could contribute to measuring how close the learning system is to stability. For this purpose, the results of successive validation samples would be compared to determine how much the system's behavior has changed and to assess whether that behavior is converging toward the behavior required to meet the Users purposes.[7]

## *What tasks must one perform to validate an HBR?*

From the broadest possible perspective, validating an HBR involves performing the following high-level tasks:

- collect as complete a set of requirements and acceptability criteria as possible
- identify the referents to be used in assessing the HBR's accuracy
- validate the developer's conceptual model for the HBR against the requirements using the referent
- analyze the conceptual model to identify areas of high complexity that can help focus later validation activities (particularly, results validation)
- validate the knowledge base against the requirements using the referent

[7] See the special topic on Validation for additional information.

- analyze the knowledge base to identify areas of high complexity that can help focus later validation activities
- validate the integrated HBR against the requirements using the referent and concentrating upon the most complex areas as suggested by the conceptual model and knowledge base complexity analysis

These tasks may seem to apply primarily to a new HBR development. However, all of the validation steps should be repeated for every new set of requirements since former validation results will depend upon the specifics of their requirements.[8] Further, any changes to an existing HBR design, knowledge base or implementation will invalidate any existing complexity analysis results. In these cases, the complexity analyses will need to be repeated, if not for the complete system, for those areas that have seen modification.

## *What information does an HBR validation plan contain?*

The plan for validating an HBR should answer the following questions:

| HBR Validation Planning Issues |
|---|
| • Have the Users articulated the requirements for the HBR and, if not, when will those requirements be available? |
| • Who will derive requirements from the User-supplied requirements and when will they make those derived requirements available? |
| • When will the acceptability criteria for the HBR be available? |
| • Who will identify the referents necessary for the HBR's validation and how much time will they need to complete that task? |
| • What purposes will SMEs serve in validating the HBR and what special knowledge and skills will they require to adequately serve that purpose? |
| • Who will identify the SMEs needed to support the HBR validation process and when will they make that information available? |
| • When will SME participation be needed in the HBR validation process? |
| • What information do the SMEs need to prepare them for their participation in the validation process? |
| • How much advance notice will SMEs need to insure their participation in the HBR validation process? |
| • Who will acquire the information to prepare the SMEs, assemble it into preparation packages and distribute it to candidate SMEs? |
| • How much time is needed to prepare SMEs for their participation? |
| • What additional arrangements must be made to facilitate SME participation and who will make those arrangements? |
| • How will the variance between multiple SMEs be characterized and controlled? |

[8] See the special topic on Requirements for additional information.

| HBR Validation Planning Issues |
|---|
| • What contingencies exist for replacing any SMEs who cannot participate as expected and what are the cost and schedule impacts of taking advantage of those contingencies? |
| • Who will collect and assemble the information needed for non-SME referents and how much time will they need to complete that task? |
| • When will the conceptual model be delivered for validation? |
| • What techniques and tools will be applied to validating the HBR's conceptual model? |
| • Where will the tools for validating the HBR's conceptual model be obtained? |
| • Who will acquire or provide the tools for HBR conceptual model validation and how much time will they need for that task? |
| • If the conceptual model validation tools need modification, who will be responsible for that task, when will that modification take place, and how much time will that require? |
| • If conceptual model validation tools need modification, what contingencies are available if the resources for their modification are not available and what impact will implementing those contingencies have upon the schedule and cost? |
| • When will the HBR's knowledge base be available for validation? |
| • How will the HBR's knowledge base be validated? |
| • How much time is available for validating the HBR's knowledge base? |
| • What techniques and tools will be applied to validating the HBR's knowledge base? |
| • Where will the tools for validating the HBR's knowledge base be obtained? |
| • Who will acquire or provide the tools for HBR knowledge base validation and how much time will they need for that task? |
| • If the knowledge base validation tools need modification, who will be responsible for task, when will that modification take place, and how much time will that require? |
| • If knowledge base validation tools need modification, what contingencies are available if the resources for their modification are not available and what impact will implementing those contingencies have upon the schedule and cost? |
| • When does the developer need input for the test plan for the integrated HBR? |
| • When does the developer need input for the scenarios used for integrated HBR testing? |
| • When does the developer plan to test the integrated HBR? |
| • When will the results from those tests be available for validation? |
| • How will the integrated HBR's test results be validated? |
| • How much time is available for validating the integrated HBR's test results? |
| • What techniques and tools will be applied to validating the integrated HBR's test results? |
| • Where will the tools for validating the integrated HBR's test results be obtained? |
| • Who will acquire or provide the tools for integrated HBR results validation and how much time will they need for that task? |
| • If the results validation tools need modification, who will be responsible for task, when will that modification take place, and how much time will that require? |
| • If results validation tools need modification, what contingencies are available if the resources for their modification are not available and what impact will implementing those contingencies have upon the schedule and cost? |

The answers to these questions not only identify the major components needed for HBR validation but will also provide the information needed to construct a preliminary validation schedule as well as identifying the events that must occur during that schedule. These questions also attempt to identify any contingencies needed should the main components of the plan fail.

Not every HBR validation effort will need to address every question. However, the V&V agent should probably examine every question individually to determine if it applies. Further, the V&V agent will need to revisit some subset of these questions with each iteration of the development process.

# HBR REQUIREMENTS

As with any other simulation, the requirements for the HBR establish the foundation necessary for validation. Validation essentially consists of comparing simulation capabilities with requirements. As a result, the credibility of any validation results cannot be stronger than the requirements upon which they depend. This section touches upon various problems associated with defining requirements for HBRs. It addresses the following questions:

- What HBR requirements should Users provide directly?
- What can one do about under-specified HBR requirements?
- What can one do about over-specified HBR requirements?
- What are acceptability criteria in the context of HBR validation?
- What does verification of HBR requirements involve?

These questions only introduce the reader to the problems associated with requirements. In many cases, HBRs share the same difficulties with requirements experienced with other simulation types. The RPG addresses these difficulties in many other sections.[9]

## *What HBR requirements should Users provide directly?*

Typically, getting complete requirements statements for any simulation from its Users can be challenging. Understanding that hurdle, at a minimum, the Users should define what jobs the humans that the corresponding HBRs represent perform, even if only at a very high level. This information should also define the levels of skill that the simulated human must represent. Much of the required cognitive capabilities for the HBR can be derived from that information. Further, if the simulation requires representing multiple

[9] See the special topic on Requirements for additional information.

humans then defining their jobs also defines their levels of cooperation in team arrangements.

In some cases, the term role may apply better than job (e.g., individuals in crowds). Beyond that, if possible, the Users should define what human effects they feel should be incorporated into the simulation. Such effects include all of the behavior moderators such as various kinds of physical and psychological stress, fatigue, emotion, motivation and various other personality differences. These effects determine how realistically the HBRs should appear in the greater simulation. Any other details that the Users care to provide will be helpful and can improve the credibility of the delivered product.

## *What can one do about under-specified HBR requirements?*

When facing under-specified requirements for an HBR one can

- derive additional requirements from the user-supplied set
- iterate with the User throughout the development process to successively refine the requirements statements

Knowing the jobs or roles that the simulated humans should perform to achieve the Users' objectives really provides a lot of information about the required HBR. From this knowledge one can identify much of the contents of the knowledge base and the cognitive skills necessary to perform the job or act in the role. This information can come from such sources as training manuals relevant to the job, skill and task testing standards (e.g., ARTEPS) and SMEs.[10]

One can also infer HBR requirements from the nature of the simulated world within which the HBR operates. The simulated objects and properties of that world define with what the HBR interacts to perform its job. Further, the other HBRs and real humans with which the HBR interacts also contribute to creating that environment. In effect, these components define the boundary conditions for the HBR and delimit the range of situations with which it must deal. It makes no sense to develop an HBR with the capabilities to reason about and respond to objects, events or situations with which it will not come in contact either through the simulated or actual worlds to which it is coupled.

A crucial part of deriving requirements is maintaining the traceability to the original User requirements. This will help to provide the necessary context when reviewing the derived requirements with the Users. It will also make propagating any User requirement changes to the derived requirements easier and more reliable.

Depending upon their accessibility, a dialogue with Users about their HBR needs can also clarify and particularize the original requirements information. This dialogue is especially necessary when the original requirements imply the need for an HBR with

[10] See the special topic on Subject Matter Experts and VV&A for additional information.

capabilities beyond what the prevailing technology can provide for the available funding and within the desired delivery schedule. Finally, as with any simulation, the Users or the Users' representative should review and approve any requirements derived for any HBR.

In some situations, the Users under specify requirements because they do not know precisely what HBR capabilities they need to accomplish their objectives. This may signify the need to adopt a simulation development paradigm that evolves requirement specificity with the development. The spiral development approach is one of the better known of these strategies. The adoption of a particular development paradigm has implications that reach much farther than validation and should be considered in this broader context. Use of evolutionary development strategies means that the validation product should evolve as the requirements become better understood. This path can also postpone much of the validation effort into the later phases of development, a risky choice that should be weighed carefully with the other program factors.

Dealing with under-specified requirements through an iterative dialogue with the User has the advantage of involving the User throughout the development process. This can greatly increase the credibility of the HBR over a process that favors simple User review of derived requirements.

## *What can one do about over-specified HBR requirements?*

Users can easily specify requirements for HBRs that exceed our current technological level. Modeling human behavior at fine-grained levels can be extremely difficult and costly. Yet, complex human behavior is commonplace in our actual world and this sophisticated behavior is often taken for granted. Further, the size of the validation effort grows exponentially as the resolution of the HBR becomes finer.

One can deal with over-specified HBR requirements by

- reducing the requirements through dialogue with the User
- augmenting the HBR capabilities with human assistance

In most cases, less capable HBRs can achieve the Users' objectives for the greater simulation. For example, a crowd need not be simulated by representing the behavior of every individual in the crowd.

In those cases where accurate fine-grained behavior is needed then this can often be achieved by augmenting the HBR with the intelligence, training and skills of an actual human role player. Here, the HBR unburdens the human role player of the monotonous and time-consuming tasks. This solution often makes use of the HBR a cost effective option.

## *What are acceptability criteria in the context of HBR validation?*

As in any type of simulation, acceptability criteria define the testable standards of functionality and performance that the HBR must meet in order to be considered adequate to achieve the Users' objectives. In other words, if an HBR meets or surpasses all acceptability criteria then it will sufficiently suit all of the User's purposes. Thus, development of the acceptability criteria is critical for the success of the HBR development. For HBRs, these should define the

- human roles that must be represented
- level of performance of those roles
- human aspects that must be represented (e.g., behavior moderators, performance limitations)

Acceptability criteria are derived from the M&S requirements. They may be either quantitative or qualitative in nature. Since SMEs can represent the Users of the HBR, they should be involved in deriving and validating the acceptability criteria. SMEs should also support the results validation process to determine if the HBR produces results that meet the acceptability criteria. This is especially true in the evaluation of qualitative criteria.

### *What does verification of HBR requirements involve?*

Verifying HBR requirements involves

- performing the same tasks associated with requirements verification for other types of models and simulations
- scoping the possible set of behaviors down to only those required to achieve User requirements
- expanding behavior requirements to identify the specific degree of detail required to serve the desired purpose

Both behavior scoping and behavior requirements expansion usually require iteration with the User and involve SME support. This effort should identify what must be validly implemented through questions that stimulate the User to defend and justify each requirement. This process often results in a more realistic set of HBR requirements and contributes to behavior scoping.

## Referents for HBR Validation

One cannot assess an HBR without one or more referents. This section examines the role of referents in HBR validation and provides some guidance for choosing them. It explores the following questions:

- What purpose does a referent for an HBR serve?
- What referents exist for validating HBRs?
- What levels of referents for HBR validation exist?
- What limitations are associated with the different referents?
- What factors contribute to choosing a referent?
- How can one choose between using SMEs or experimental data as referents for HBR validation?

Referents, like requirements, contribute knowledge essential for validating any type of simulation, HBR or not. Thus, a reader desiring more information on referents should consult the other parts of the RPG that discuss referents.

## *What purpose does a referent for an HBR serve?*

A simulation's referent defines the standard against which to measure its representational capabilities to determine the accuracy of its representations. A simulation's accuracy or error from reality cannot be assessed without a referent. In an HBR, the referent defines what level of human behavior is being represented.

## *What referents exist for validating HBRs?*

A referent represents the total collection of knowledge about a particular subject, in this case, human behavior under various circumstances. Referents for HBRs can come from

- SMEs[11]
- empirical observations or experimental data from actual operations
- validated models of various aspects of human behavior
- validated models of the physiological processes underlying human behavior
- validated models of sociological phenomena (useful particularly for modeling groups of people)
- validated simulations of human behavior

By far the most employed referents for validating HBRs come from SMEs. They provide information most easily interpreted. However, several pitfalls exist in employing SMEs. Most notably, SMEs filter the information they provide based upon their internal models of the HBR and the situation to which it responds. This filtering introduces variability in SME opinions that can be difficult to assess and reconcile.

[11] See the special topic on Subject Matter Experts and VV&A for additional information.

Empirical data provide a wealth of information of such factors as response time under various conditions. Models of sociological phenomena apply primarily to representations of groups of humans.

The degree to which a referent is useful depends upon the

- specific phenomena represented by the HBR
- particular purposes for the HBR
- expertise available for evaluating and applying the referents

For example, a psychologist might be necessary to interpret the validity constraints of any model of psychological processes chosen as a referent.[12]

## *What levels of referents for HBR validation exist?*

Validating the models of a simulation involves comparing the characteristics of those model abstractions with the referents available. This comparison identifies where the models coincide with the referents and where they deviate from them. This information can then help to determine how well the models serve specific purposes. Referents include expert opinions, experimental observations and theoretical approximations of the modeled phenomena. These referents represent the best obtainable knowledge of the phenomena of interest.

Like models of complex physical processes, HBR models can be validated at many different levels of abstraction. The figure below illustrates six levels of model correspondence for HBRs. In effect, these levels define six levels of referents. These referent levels provide the standards against which one can compare an HBR's behavior to determine, or test, its correspondence with the referent.[13]

| Levels of HBR Correspondence and Referents |
|---|
| • Domain Correspondence |
| • Sociological Correspondence |
| • Psychological Correspondence |
| • Physiological Correspondence |
| • Computational Correspondence |
| • Physical Correspondence |

### Domain Correspondence

[12] See the special topic on Validation for additional information.
[13] See the reference document on HBR Literature Review for additional information.

Domain experts, also known as SMEs,[14] know what results typical human behavior in their particular domains should produce. This knowledge permits those experts to examine HBR knowledge bases, observe HBR performance in their domains, and assess, often qualitatively, how realistically those models represent the necessary human behavior for a purpose. In this process, the SMEs themselves become the referents for the purpose. Domain correspondence testing was first widely employed in expert system construction (i.e., empirical techniques). Some have called domain correspondence testing the ***Turing test*** although most of the recent Turing test interpretations differ significantly from what Alan Turing originally intended as a sufficiency test of machine intelligence.[15]

Domain referents can also come from experimental data. Some quantitative experimental data exists on actual human performance in various battlefield situations (e.g., data from instrumented ranges) and from humans performing very specific cognitive tasks under controlled conditions. The results produced by HBRs can be compared with this data rather than compared against the opinions of SMEs. This performance data, when available, strengthens the results of domain correspondence testing. Regrettably, much domain-specific experimental data is very sparse and applies only to narrow situations. Often the experimental conditions for data collection are very poorly controlled and characterized. These source data problems weaken any validation done against them. However, as the technology develops, better experimental data for different domains will improve and become more widely accessible.

## Sociological Correspondence

HBRs that replicate groups of interacting people should possess sociological correspondence. This includes disordered groups such as crowds as well as groups operating within some organizational structure. As with psychological validation, a rich body of sociological knowledge exists from which referents and tests for sociological correspondence can be drawn. This knowledge includes both models describing sociological phenomena and experimental observations. Sociological experiments also provide well-established experimental protocols to support the design of sociological correspondence tests.

Sociological correspondence testing is similar to psychological correspondence testing. Sociological correspondence can also be tested against the observations of psychology and sociology professionals that act as SMEs. These include experts in organizational structure and dynamics. Current sociological knowledge permits the testing of behavior manifested by groups as well as of the interaction dynamics between the members of those groups. Testing sociological validity is particularly important with simulations of human groups cooperating to perform some task and may not be necessary when representing the actions of an isolated individual.

[14] See the special topic on Subject Matter Experts and VV&A for additional information.
[15] See the reference document on V&V Techniques for additional information.

## Psychological Correspondence

A vast body of knowledge exists about human psychology. This knowledge includes numerous abstract models of many different aspects of human behavior as well as an enormous volume of published experimental data on actual human performance under different circumstances of interest that validate these theories to some degree. This knowledge enables the testing of another form of correspondence between HBRs and reality.

Testing the psychological correspondence of an HBR starts with identifying the psychological models and experimental data appropriate for the problem domain. Both models and experimental data establish the referents against which to compare HBR behavior. Experimental data can completely establish a referent or augment that created by psychological models. Then, a set of carefully controlled experiments produces the data on the HBR performance. Comparing these data against the referents validates the models underlying the HBR of an individual.

Testing psychological correspondence creates stronger validation than domain correspondence testing alone because of its linkage to the underlying psychological phenomena. This linkage to the founding phenomenology means that the entire problem space need not be explored as in testing domain correspondence because the psychological models inherently represent the nonlinearities associated with the behavior they generate. Further, the experimental data from which referents are drawn has more likely been obtained under carefully controlled experimental conditions and is therefore more repeatable than that obtained from domain-specific experiments. Psychological correspondence testing enables validation of all of the HBR model components both as separate functions and as an integrated whole.

## Physiological Correspondence

A considerable collection of experimental data and, recently developed, verifiable theory of neurophysiological processes begin to establish additional referents against which to compare HBR performance. HBRs have one significant advantage over the actual physiological systems from which this data originates; their detailed workings are easier to directly observe. Simulations that have physiological correspondence more likely behave like real people especially under conditions where non-neurological physiology contributes (e.g., fatigue and injury). This sort of evaluation is much closer to what has traditionally been done to validate physical system representations (i.e., non-human systems).

In the past, this kind of validation has been difficult because the physiology of the human nervous system was not understood well enough to correlate physiological observations with cognitive behavior except at extremely low levels (e.g., primitive vision). However, recent advances in noninvasive measurement techniques (e.g., MRI, PET) have improved our understanding of the linkage between cognitive behavior and physiological observations and created a large repository of potential validation data.

As this area of experiment improves, comparing these experimental results with HBR designs and performance will become easier and more meaningful. This form of correspondence testing might be particularly well suited for validating behavior moderators, such as stress and emotion, and integrated models of human behavior that incorporate such effects behavior moderators.

## Computational Correspondence

The human nervous system replicates of a large collection of interacting computational devices through which the brain performs well-understood computational tasks (e.g., receiving input, storing information, making decisions, producing output). While the anatomical and physiological specifics of the brain's computational processes are not currently well understood, the fact that the brain performs these functions is indisputable. All human brains have limited input, output and computational bandwidths and storage capabilities.

Data from psychological experiments provide some bandwidth and storage limit referents for various classes of tasks. These data, together with a small amount of theory on the brain's computational performance, establish a referent for computational correspondence testing. This level of correspondence testing is reasonably easy to perform. While tests of computational correspondence are not alone sufficient to validate HBRs for most purposes of interest, this form of validation complements other validation tests by describing necessary limits upon realistic behavior. Computational correspondence testing provides a means to validate important human performance limitations. As computational correspondence testing becomes more widely applied, the breadth and depth of the referents will improve and these advances will improve the quality of validation these tests can provide.

## Physical Correspondence

Basic physical laws limit the performance of humans just as they do in any strictly physical system (i.e., not a system whose behavior depends upon knowledge and goals). Consequently, the performance of an HBR can be compared against the limits predicted by the physics governing the human's activities. Representations that exceed these limits inaccurately predict the behavior manifested by the human. In fact, many aspects of observable human performance may originate from these reasonably simple physical limitations.

Much about the physics of real neural systems, the most enigmatic component of humans, remains a relatively unexplored area and so lacks the theory and experimental results to create reliable referents. The physical limitations represented by models and simulations have been extensively validated for non-human systems but not widely applied, as yet, to HBRs, especially to the nervous system functions. Nevertheless, humans are physical devices constrained by physical laws and so correspondence should exist between HBRs and the underlying physics for them to produce accurate results.

Realizing that humans are subject to the same limitations imposed upon all physical systems can significantly reduce the inherent complexity of validating HBRs at a physical level. Significant research must still be done to make this level of validation useful for largely cognitive functions and, like computational correspondence testing, physical correspondence testing may never provide the grounds for sufficient validation alone. However, physical correspondence does provide one more test to help guarantee the accuracy of HBR performance.

### Correspondence and Validity

Several HBRs have been developed and validated for a variety of purposes. This validation can be described in terms of model correspondence testing. A recent National Research Council study provides an excellent survey of HBR technology [Pew and Mavor, 1998]. These study results discuss the validation of many of the existing and developing HBRs.

In all, six levels of correspondence establish the validity of HBRs. An HBR that has correspondence at all six levels best approximates human behavior for all purposes. Most purposes may only require correspondence in one or two of these areas or over four areas under very specific conditions. These choices can appreciably limit the complexity, cost and risk of the HBR validation process. The theoretical models and experimental data associated with the psychological, sociological, physiological, computational and physical levels can tremendously reduce the need to search the entire problem space of the intended domain during validation. Validation of any HBR with widely accepted theory or data sets could produce a system that need not be extensively revalidated for each new purpose. This will drastically reduce the cost of validating existing HBRs and hasten their widespread employment.

## *What limitations are associated with the different referents?*

All of the techniques applied to validating existing HBRs have significant limitations.[16] As mentioned, testing domain correspondence requires unrealistic searches of very large and nonlinear behavior spaces. Testing psychological and physiological correspondences requires extensive validated models of psychological and physiological phenomena. While many comprehensive psychological models exist, relatively few of them have been applied to HBR validation, especially for simulation purposes.

Like the physiological models, many psychological models deal with very restricted behavior spaces. These limitations prevent their useful application to HBRs representing behavior for realistic situations. As psychological and physiological models become richer and more consistent, their utility for HBR validation will increase. As with models and simulations of physical systems, model correspondence testing

[16] See the reference document on V&V Techniques for additional information.

should be done at several levels of abstraction. Only consistent results between these different levels can guarantee validity. At this point, and probably forever, no single level of correspondence testing should be sufficient for any purpose.

## *What factors contribute to choosing a referent?*

A referent must satisfy at least three basic criteria:

- **It must be available**. This criterion implies that HBR referents come primarily from our collected knowledge of human behavior. The exception to this is the referent against which the accuracy of the HBR's perceptions will be assessed. In this case, the referent comes from the simulated world, the ground truth for the HBR's perceptions.
- **The Users must consider it credible**. Since any assessment of an HBR's accuracy depends upon the degree to which the referent reflects the real humans being simulated and accuracy is an important representational property, the Users must believe in the referent to see any assessments of the HBR's accuracy as credible.
- **The referent must contain the same level of detail as the HBR**. Choosing a referent with a higher degree of abstraction than the HBR represents will diminish that referent's value for assessing accuracy. Choosing a referent with more detail than the simulation can represent wastes valuable resources for collecting and compiling the referent data and may increase the amount of effort expended for validation.

## *How can one choose between using SMEs or experimental data as referents for HBR validation?*

The choice to use an SME over experimental data depends entirely upon the appropriateness of the SME's experience to the situation being modeled and upon the richness and coincidence with the situations of interest of the experimental data. Experimental data, if properly collected and interpreted, provides insight unbiased by the personal preference often present in SMEs. However, the applicability of experimental data to a particular purpose is usually severely limited by the

- size of the data set
- extent of the controls exerted upon the conditions under which it was collected
- uncertainties associated with the experimental conditions, measurement, and analysis

SMEs generally offer a broader but more qualitative source of referents. SMEs perform particularly well when the situation closely resembles their experience and education. However, people are very powerful generalizers. This tends to broaden the validity of

SME guidance over circumstances that do not fall squarely within their direct experience. Nevertheless, care should be taken because real humans often invent an answer when they only know something resembling the question at hand. This tendency can lead to incorrect advice and invalid HBR behavior being judged either correct or incorrect when it is not.

Additional information on the use of SMEs in validation can be found in the special topic on SMEs.

# Subject Matter Experts For HBR Validation

Subject matter experts (SMEs) serve many important purposes in HBR validation. But, their use also brings its own problems. This section considers many of the issues associated with using SMEs in HBR validation. It does this in the following questions:

- How can SMEs support HBR validation?
- What factors constrain the use of SMEs for HBR validation?
- Where does SME involvement in HBR validation matter the most?
- How should one select the SMEs to participate in HBR validation?
- How should one prepare SMEs for HBR validation tasks?
- What documentation should result from SME involvement in HBR validation?
- What problems can occur when using itinerant SMEs for HBR validation?
- How can one address the problems caused by using itinerant SMEs?
- How should one handle the need to use SMEs who have no direct HBR experience?
- Where should SMEs for HBR validation fit organizationally?
- What tasks should an SME-Engineer liaison perform?

Again, employing SMEs in the validation process is not unique to HBRs. The RPG addresses many of the general issues associated with SMEs in other places.[17]

## *How can SMEs support HBR validation?*

SMEs play a crucial role in HBR validation in every phase of the validation process because of the inherent complexity of HBRs and the human behavior they strive to represent. Optimally, SMEs work with the developer and the V&V agent throughout the development process, from requirements verification through results validation. SMEs serve three essential purposes in HBR validation:

[17] See the special topic on Subject Matter Experts and VV&A for additional information.

- They supply referent knowledge through their training, education and experience
- They compare HBR capabilities against their referent knowledge to assess the accuracy of those capabilities
- They assess the degree to which that accuracy meets the User requirements.

SMEs, usually highly trained individuals with well-developed areas of expertise, can perform many specific tasks relevant to HBR validation such as those listed in the table below.

| Typical SME Tasks in HBR Validation |
|---|
| • Determine if the collection of stated and derived HBR requirements adequately represent the User's need |
| • Help choose additional referents against which the HBR can be compared for accuracy |
| • Support the validation of the HBR conceptual model |
| • Assess the correctness, completeness and consistency of the HBR knowledge base |
| • Help with incrementally checking the correctness of the HBR design and implementation |
| • Support the selection and design of the scenarios for HBR testing |
| • Assess the correctness and consistency of the test scenarios in representing the User's purpose |
| • Assist the design of the HBR testing procedures to assure that they introduce no unrealistic artifacts into the test results and that they adequately represent the User's purposes for the HBR |
| • Guide the HBR testing process to assure the accurate execution of the testing procedures and test scenarios and to better understand the results obtained |
| • Analyze the HBR test results to assess the validity of the final product for the User's purposes |

The jobs an SME can perform in HBR validation strongly depend upon the SMEs abilities and availability to the developer and V&V agent.[18]

## *What factors constrain the use of SMEs for HBR validation?*

Ideally, SMEs will be employed throughout the HBR validation process. However, several factors can limit this ideal:

- **Expertise** -- An SME with limited expertise can only be usefully employed where the HBR represents that knowledge. Further, only SMEs with the right expertise may understand the form of the information used and produced by the HBR

[18] See the special topic on Subject Matter Experts and VV&A for additional information.

(e.g., production rules vs. neural network weights). Using SMEs where they are not qualified can only give a false sense of security and may lead to inappropriate accreditation.

- **Availability** -- SMEs may have limited availability to the developer and V&V agent. This is especially true for those SMEs with considerable experience and, therefore, utility to their operational unit. Often the best SMEs are also the least available resources.
- **Funding** -- Many HBR development projects have mistakenly assumed that the funding for SME involvement will come from User organizations later to find those organizations only begrudgingly staff SME requirements to a minimum level. Funding from outside the User organization can often improve SME expertise and availability.
- **Time** -- The pace and schedule of the HBR development effort should match SME availability schedules. Sufficient time should be allotted for their participation. SMEs also require time to become familiar with the project, the User's detailed requirements and acceptability criteria, their roles in the project, and the HBR itself as well as to perform their assessments. Failure to allocate sufficient time in the schedule for all aspects of SME involvement will limit their utility and effectiveness.

If a separate reasoning engine is being developed for a purpose, then the SME's participation is useful only during the development of the knowledge base and later during integrated testing. Several languages exist for representing expert knowledge. In some cases, these directly enable the SME to encode the desired expert knowledge. In other cases, these knowledge representation languages are complex programming languages and require a trained knowledge or software engineer to build the knowledge base (e.g., SOAR). In these cases, the SME should work closely with the engineer to build and validate the knowledge base.

In yet other cases, the mechanisms for encoding and executing expert knowledge are tightly coupled and cryptic (e.g., systems using genetic algorithms and neural networks), providing the SME with only observable output of the behavior system. In these cases, the SME provides expert knowledge to the developing engineers but must devote careful attention to exhaustive testing of the system behaviors. Of course, in these situations, the development process relies on several iterations to get the system behavior right. Even though the SME may be integrally involved in the creation of the knowledge base, the entire system (i.e., knowledge base integrated with the underlying inference engine) should be validated since different inference engines using the same knowledge bases can produce different results.

## *Where does SME involvement in HBR validation matter the most?*

SMEs make their most important contributions to the

- development and validation of acceptability criteria

- validation of the HBR conceptual model[19]
- results validation process

If SMEs have limited availability then one should apply them to these three phases of the validation process. Integrating SMEs tightly into the validation process provides better validation results. [20]

## *How should one select the SMEs to participate in HBR validation?*

The selection of appropriate SMEs is crucial to successfully employing them in any HBR validation effort. Selection methods and typical SME challenges involving cross-domain knowledge transfer and the psychology of picking SMEs and extracting information from them are covered in the special topic on SMEs.[21]

## *How should one prepare SMEs for HBR validation tasks?*

The following information should be provided to all SMEs prior to their involvement in any HBR validation task:

| Preparatory Information for SMEs |
|---|
| • User's objectives |
| • Stated requirements |
| • Derived requirements |
| • Scenarios |
| • Conceptual model representation |
| • HBR modeling approach and design |
| • Knowledge base representation |
| • Guidance delivered by other SMEs |
| • Test plan and procedures |
| • Testing conditions, controls, scenario, problems encountered and test data |
| • Past validation conclusions and changes made to the HBR since then |

Clearly, the stage of the validation process determines how much the SME should be prepared and what information they need. The list above applies to the most general situation. In addition, SMEs should be familiarized with the use of any tools or special data formats to which they will be exposed (e.g., behavior audit trails).

[19] See the special topic on Conceptual Model Development and Validation for additional information.
[20] See the special topic on Subject Matter Experts and VV&A for additional information.
[21] See the special topic on Subject Matter Experts and VV&A for additional information.

## *What documentation should result from SME involvement in HBR validation?*

In all phases, each review of each HBR instance by SMEs should be documented. This documentation should include

- who was doing the review
- instructions given to the reviewer
- subjects reviewed
- test plans exercised and their results
- decisions made or conclusions drawn

These documents contribute to the body of evidence supporting HBR validation for that particular purpose. This documentation is important both for accreditation for the immediate purpose and for later reuse of the HBR for other purposes.[22]

## *What problems can occur when using itinerant SMEs for HBR validation?*

Many HBR development efforts have had to face using SMEs with limited continuous availability. As a result, their access to a particular SME may be infrequent and they may only be able to meet their SME needs by using a series of SMEs assigned to them on an as needed basis. This common situation constitutes the itinerant SME problem. Use of itinerant SMEs may occur because of personnel rotation in the User organization or reassignment to other tasks more important to that organization. The use of itinerant SMEs can have significant impact upon the validation effort:

- each new SME will have no or only limited understanding of the HBR's specific purpose, requirements, functions and modeling issues
- different SMEs will provide different expert guidance to V&V and development personnel for the same circumstances due to individual differences in applying the same doctrine[23]
- an SME will have to become acquainted or reacquainted with the state of the validation effort and the development project with each visit, taking up time that would normally be used for the validation process
- new SMEs may revisit issues that were resolved in the past due to individual differences
- the level and specialty of SME expertise may vary wildly from one person to another

[22] See the special topic on Subject Matter Experts and VV&A for additional information.
[23] This consistency issue almost always occurs in using itinerate SMEs.

- SME enthusiasm and interest in the validation and development efforts may be minimal because they have no stake or lasting obligation to the effort

SMEs consider themselves experts in their particular domains. The impreciseness of most domains to which HBRs apply creates the opportunity for diverse and seemingly inconsistent expert opinions. This occurs because of the complexity associated with developing HBRs rather than because of problems with the SMEs or their knowledge. Therefore, this problem can only be ameliorated and tolerated.

## *How can one address the problems caused by using itinerant SMEs?*

Appointing a member of the V&V or development teams as the SME-Engineer liaison can ameliorate many of the problems associated with using itinerant SMEs. This person should have experience that enables him to talk effectively and efficiently with both SMEs and technical personnel. This person must understand the HBR's requirements as well as the technical issues associated with its development and testing. Further, this person should be a permanent part of the V&V or development team to provide the needed continuity between itinerant SMEs.

## *How should one handle the need to use SMEs who have no direct HBR experience?*

SMEs with prior experience in HBR development and validation can simplify their management. However, these people may not be available. SMEs often come to HBR development projects with little or no prior experience in this specialized modeling area. When this occurs, appointing an SME-Engineer liaison from the V&V or development team may offer the only realistic solution.

## *Where should SMEs for HBR validation fit organizationally?*

Ideally, the required SMEs will be attached to the V&V or development team throughout the development and testing of an HBR. The SMEs should report to either the V&V agent or the accreditation agent, whichever is most appropriate. This organization assures continuous SME access and integrates the SMEs with the development team.

The dedication of SMEs to an HBR development effort also deepens SME involvement and significantly improves their effectiveness, efficiency and consistency by reducing their need to reacquire system and project specific expertise with each different exposure. This same philosophy applies to efforts modifying existing HBRs for specific purposes. In these cases, continuous SME involvement is especially important since most changes to HBR behavior involve iterative modification of their knowledge bases. Here, the change-testing cycle can occur several times per day with each cycle requiring continuous SME involvement. More commonly, SMEs are assigned temporarily to support the validation or development effort. This situation brings special problems and management requirements.

### *What tasks should an SME-Engineer liaison perform?*

The person performing as an HBR development effort's SME-engineer liaison should

- Help the SME become familiar with the HBR's purpose, requirements, general technical approach and modeling issues before they interact with the V&V team or development personnel
- Participate in all interactions between SMEs and the technical staff to become the institutional memory of previous SME activities
- Resolve conflicts between the tactically sound but different approaches to military operational questions and situations that often occur when different SMEs are employed
- Maintain the consistency between the guidance offered by different SMEs
- Guide the interactions between the SME and technical personnel to avoid issues resolved by past interactions and optimize covering new ground
- Translate terms and concepts between the user domain and the development domain
- Insure that SMEs with different specialties are applied to problems within their areas of expertise
- Assure that the functionality and fidelity prescribed by the SME for the User's purpose remain within the scope of the effort's requirements

In general, the SME-Engineer liaison efficiently couples the technical development people with the SME domain and resolves any problems that might occur because of the different conceptual perspectives.

## HBR Conceptual Model Validation

Conceptual model validation delivers the first glimpse into an HBR's validity and serves a crucial purpose in building credibility with the User. Historically, few HBRs development efforts have constructed and used conceptual models but, in doing so, they have failed to take advantage of their power. The guidance herein strongly recommends building and using conceptual models for HBR validation. This section addresses the questions associated with conceptual models specifically for validating HBRs. It contains the following questions:

- What information can the conceptual model provide about the HBR's complexity?
- What tools exist to support validating an HBR conceptual model?

Since conceptual models form an essential part of the validation strategy proposed in the RPG for any type of simulation, the RPG contains much more information about conceptual models in other places. This information largely applies to HBRs as well.

## *What purposes does a conceptual model serve for HBR validation?*

As the first product of the development effort, an HBR's conceptual model stands as the first point for validation in a new development. It provides the triage point in using a legacy HBR that will determine how much modification the HBR will need to achieve new objectives. Early validation can help to reduce development or modification risk by uncovering representational problems and shortcomings before significant investments have been made.

An HBR's conceptual model also supplies considerable information about the complexity of the HBR and where that complexity lies in the HBR's behavior space. This information can help to tailor subsequent, more labor intensive, validation steps and to focus validation effort upon those elements of the HBR mechanisms and knowledge base where complexity is greatest. This contribution can greatly reduce the cost of knowledge base and results validation as well as improving the credibility of their products.

## *What should the conceptual model for an HBR generally contain?*

The conceptual model describes the developer's interpretation of what is needed to achieve the User's objectives. As a result, the conceptual model for a simulation that involves HBR should define the

- tasks the HBR must perform
- objects and properties of those objects that the HBR can sense
- objects and properties of those objects that the HBR can explicitly change through its actions
- situations that identify when the HBR must change its task mix
- goals associated with each task
- situations that identify when the HBR must change its goal mix within each task
- effects of the internal factors that can moderate the HBR's responses (e.g., fatigue, injury, fear)
- knowledge that the HBR must possess to manifest the proper responses to the proper situations

If HBRs (or the same HBR infrastructure) supports the simulation of several different people then the conceptual model should also include information that identifies the

- different types of people to be simulated and the roles they perform
- differences in the knowledge that each person possesses
- knowledge that the simulated people rely upon to facilitate their interactions
- personality factors that distinguish one person from another

The situations and responses should include those required to support interactions between simulated people and between simulated people and real people.[24]

## *How should one validate an HBR conceptual model?*

In general, validating HBR conceptual models should compare selected component models with current knowledge about human behavior to assess their validity. To this end, SMEs, familiar with the User's objectives and the scenarios, should examine that part of the conceptual model describing the HBR to assure its completeness and consistency with user doctrine. In the context of these objectives and the scenarios, they should determine if the

- different types of people and groups simulated are sufficient to address the different types required to achieve the purpose
- sets of situations and responses are sufficient to accommodate the scenarios required to achieve the purpose
- responses to each situation are adequately realistic
- influences of behavior moderators are adequately represented
- knowledge possessed by each type of simulated person is sufficient to create adequately realistic behavior

The SME may require assistance from the development team in reading and understanding the conceptual model representation (e.g., Unified Modeling Language (UML) may need to be translated for SMEs) if they are not already familiar with the language or format in which it is recorded.

HBR conceptual model validation is iterative with the Developer responding to SME guidance by correcting the conceptual model then letting the SMEs reexamine the modifications. This process is also closely tied to HBR requirements verification and acceptability criteria development.[25]

[24] See the special topic on Conceptual Model Development and Validation for additional information.
[25] See the special topic on Conceptual Model Development and Validation for additional information.

### *What information can the conceptual model provide about the HBR's complexity?*

Many factors contribute to HBR complexity, essentially the size and shape of its behavior space. A well formulated conceptual model can identify such factors as the

- approximate number, ranges and intersections of sensory inputs
- approximate number, ranges and intersections of action outputs
- approximate number and intersections of different situations to which the HBR must respond
- conditions that distinguish one situation from another
- number and sophistication of behavior moderators modeled
- aspects of behavior represented by stochastic models

Each of these factors contributes to the complexity of the validation problem and knowledge of which parts of the behavior space they influence identifies where to focus the results validation effort.

### *What tools exist to support validating an HBR conceptual model?*

Unfortunately, no tools exist that specifically support the validation of HBR conceptual models. However, some general purpose modeling tools may be applicable and may even be used by the developer. For example, the developer may have used a Universal Modeling Language (UML) tool to construct the conceptual model. The developer will have further information on these and can assist in their use to support the validation effort.

## HBR Knowledge Base Validation

Knowledge bases contribute to making HBRs unique among simulations. They also provide another vista from which to gain an early look into HBR validity and their validation contributes a valuable piece to the credibility puzzle. This section examines the issues associated with validating an HBR's knowledge base in the following questions:

- How does the HBR's knowledge base contribute to the validity of the HBR's behavior?
- How does the HBR's knowledge representation affect validating its knowledge base?

- How can validating the HBR's knowledge base contribute to its results validation?
- What techniques and tools exist to support validating HBR knowledge bases?
- What is KBS VVE&T technology and how mature is it?
- What does KBS VVE&T technology offer for HBR knowledge base validation?

HBR knowledge bases can contain an immense amount of information and associated complexity. Fortunately, HBR knowledge bases can share many of the properties that expert system knowledge bases possess. This similarity avails the considerable technology base for validating expert system knowledge bases to HBR validation.

## *How does the HBR's knowledge base contribute to the validity of the HBR's behavior?*

The HBR's knowledge base essentially contains the computer program that determines the HBR's response to the stimuli it receives from the simulated world. At a minimum, the knowledge base largely determines the HBR's cognitive behavior. It may also contribute to the manifestations of emotion upon behavior. Thus, an invalid knowledge base will likely generate invalid behavior. However, the validity of the knowledge base cannot guarantee the validity of the HBR since many other components contribute to the overall validity (e.g., behavior engine, sensor and effector models, behavior moderator models).

## *How does the HBR's knowledge representation affect validating its knowledge base?*

The difficulty and effectiveness of knowledge base validation depends strongly upon the

- knowledge representation used
- validation methods employed
- tools chosen to support validation

The selection of the knowledge representation affects all of these factors.

Some knowledge representations lend themselves to nearly direct human scrutiny. One of the most popular of these accessible representations is the production rule (i.e., IF (some condition is true) THEN (take some action)). In most cases, these representations can be read and easily interpreted by SMEs. Other representations such as semantic nets, probabilistic inference networks and decision tables, while somewhat less accessible than production rules, can easily be translated into accessible form automatically. Such representations as neural networks and genetic algorithms are among the least accessible although some promising work has been done to make their machine translation easier for a small subset of variants.

SMEs, the most common validation method, often require the assistance of knowledge engineers even for such accessible representations as production rules. Use of less accessible knowledge representations can significantly complicate the SME's job and can often obscure the underlying causes of anomalous behavior thereby making direct validation extremely difficult and unreliable. Choosing less accessible representations places more reliance upon tools to translate the meanings of the knowledge elements into understandable forms. In some cases, automated tools exist or are being developed to directly aid the validation of knowledge bases using less accessible representations. However, most of these tools are experimental and are not widely employed in HBRs or supported for that purpose.

Tools, where they exist, can significantly increase the reliability and decrease the human labor required to validate knowledge bases. Ironically, the largest selection of applicable theory, techniques and tools for knowledge base validation exist for production rules, the most accessible representation. However, tools are currently being developed to support a wide range of representation options so some research into this field could reap great benefits.

## *How can validating the HBR's knowledge base contribute to its results validation?*

Validating a knowledge base gives detailed insight into the knowledge that the HBR maintains and the structure of that knowledge. This insight can tell not only how valid the knowledge is for a particular purpose but also where the complexities of that knowledge probably lie. This information can help to focus the results validation activity towards those areas where the HBR's behavior is most complex and difficult to characterize reliably.

## *What techniques and tools exist to support validating HBR knowledge bases?*

Knowledge base validation should compare the selected knowledge base components with current knowledge about human behavior to assess their validity. Relatively little technology explicitly exists for this purpose. However, significant resources exist to support verifying, validating, evaluating and testing (VVE&T) knowledge-based systems (KBSs). While most of these resources were developed primarily for expert systems (largely for medical applications), much of this technology has direct application to HBR knowledge bases. A search of the literature on VVE&T of KBSs [Harmon, 1998] identified several resources that provide good overviews of this active research area. These resources include books [Adelman, 1991; Ayel and Laurent, 1991; Gupta, 1992; Smith and Kandel] and survey articles [Geissman and Schultz, 1988; Grogono and Baatarekh, et al. 1992; Grogono and Preece, et al., 1992; Hamilton and Kelly, 1991; Lopez, et al., 1990; Lydiard, 1992; Mengshoel and Delab, 1993; O'Keefe and O'Leary, 1993; Preece and Suen, 1993].

## *What is KBS VVE&T technology and how mature is it?*

Significant investment has gone into the development of KBSs to perform a variety of expert functions including diagnosis, decision support and automatic control. KBSs have been applied to such critical applications as flight control, financial management, disease diagnosis, and treatment recommendation. These applications, together with the enormous success of augmenting human expertise with machine intelligence, have driven the development of technology to measure and improve the validity of KBS behavior. HBRs are a class of KBS so the technology supporting the VVE&T of KBSs, represents by its theory, techniques, tools and accumulated experience, presents a tremendous and, heretofore, untapped resource for HBR validation.

The theory relevant to the VVE&T of KBSs seems comparatively small but it covers all of the important problems. In addition, developing the theory that underlies the behavior of KBSs is a challenging task that has only recently seen some promising advances. Until a comprehensive and consistent theory of intelligent systems exists, the theory supporting VVE&T of KBSs will likely remain as loosely coupled conceptual islands. On the other hand, a myriad of verification and validation (V&V) techniques and tools have been proposed, developed and tested for KBSs. The tools vary from single tools with limited capabilities and associated with specific expert systems to rich integrated tool sets that apply to any KBS written in a particular programming language (e.g., PROLOG or OPS-5) or using a specific expert system shell. An enormous amount of experience in VVE&T of KBSs exists for diverse purposes. By far, most of this experience is related to medical applications where the results from any KBS can have life threatening consequences.

KBS VVE&T technology is relatively well developed (at least, when compared with the technology for validating HBRs. Many of its tools could be applied to future HBRs with appropriately selected implementation strategies. Most of the KBS VVE&T developments specifically address KBSs using production rule knowledge representations. This focus upon rule representations currently has limited impact on the validation of HBRs since they also commonly use production rule representations. None of the existing HBRs were designed to take advantage of any of the existing KBS VVE&T tools so modification of either the HBRs or the tools will be necessary to directly benefit from these existing resources. Further, all of the VVE&T theory, techniques and tools apply only to the cognitive functions of HBRs and cannot be used for validation of the effects of behavior moderators or performance limitations.

## *What does KBS VVE&T technology offer for HBR knowledge base validation?*

Essentially, all HBRs fall within a specialized class of knowledge-based system (KBS). Many problems in verifying, validating, evaluating and testing knowledge-based systems have already been explored, some extensively, and most of these problems apply directly to HBRs. Several tools and techniques already exist and may be applicable to HBRs. These techniques and tools can help assess the

completeness/coverage, consistency/coherence, and redundancy of the knowledge bases.

Tools and techniques have been developed for reasoning paradigms such as nonmonotonic logic, case-based reasoning, tabular knowledge representation, equations, weighted production rules, meta-knowledge and dynamic object properties. Considerable experience exists in verifying and validating KBSs and in applying these tools and techniques for a variety of different problems including automatic refinement, knowledge base verification, and evaluation of subjective criteria, large knowledge bases, and diverse application domains. Becoming familiar with what is available in this area that can be adapted to address different aspects of HBR validation can reduce the need, expense and time of inventing new techniques for validating models and simulations of human behavior.

Designers of the HBRs should look at the design tradeoffs in employing or adapting an existing KBS with accompanying tools for verifying and validating the knowledge bases. Rule-based systems have, by far, the most developed theory, tools and experience, so the option of using rule-based or production representations of knowledge deserves special consideration. Several expert system shells or frameworks have been developed that can be applied to HBR development and some of these have extensive tools for verification and validation associated with them.

If no existing KBS frameworks can meet the HBR requirements then, perhaps, an existing artificial intelligence programming language (e.g., OPS-5, Prolog) may work. Some tools have been developed for verifying and validating systems developed in these languages. While choosing to develop an HBR from scratch poses the most exciting option it is the least desirable choice from a validation perspective because no tools will exist to support validation. That option ignores over 30 years of intensive development and lessons learned from artificial intelligence and cognitive science. This choice also prevents from capitalizing on 27 years of develop in verification and validation of KBSs.

When facing validating a legacy system, searching for applicable KBS techniques and tools could provide some assistance. The HBR developer may have created tools and techniques for their use that could benefit the V&V agent directly. Tools and techniques may exist which may be adapted with relatively little effort, thus reducing the amount of validation effort required and improving the quality of the effort expended. Further, if the legacy HBR was implemented in an artificial intelligence programming language for which verification and validation tools exist, then these tools may be helpful with little or no adaptation. As a last resort, a search of the existing KBS literature [Harmon, 1998] will probably provide some guidance for solving specific human behavior representation V&V problems (e.g., knowledge base coherence, completeness, redundancy, large knowledge bases, representations of fuzzy knowledge, nonmonotonic knowledge, uncertainty, incomplete knowledge).

# HBR Results Validation

Results validation takes the final essential step in building the case for HBR validity.[26] No User should accept an HBR without some form of sufficient results validation. Unfortunately, integrated HBRs can manifest extremely complex behavior over immense state spaces and, thus, make their complete validation impractical or infeasible. Fortunately, the previous steps of validating the conceptual model and knowledge base provide information essential for guiding the results validation process so as to make it cost effective. This section explores the issues of HBR results validation in the following questions:

- What does results validation mean for HBRs?
- How important is results validation of HBRs?
- What makes HBR results validation so difficult?
- How should one test to discover an HBR's validity?
- What purposes do SMEs play in HBR results validation?
- What does behavior scoping involve and what purposes does it serve?
- How can problems discovered during HBR testing be resolved?
- How can one focus HBR results validation to improve its cost effectiveness?
- What purpose does a behavior auditing capability serve in HBR validation?
- How can validating an HBR's conceptual model help its results validation?
- How can separately validating an HBR's knowledge base help its results validation?
- How do interoperability issues affect the validity of HBR results?
- What can one do when validation activities find HBR behavior anomalies?

HBR results validation involves all of the other parts of the simulation system. This only adds to the complexity faced in this process. Problems in other parts of the simulated representation may obscure or create problems in the HBR. For this reason, one should test the HBR within a validated simulated world if at all practical.

## *What does results validation mean for HBRs?*

Results validation evaluates an HBR integrated with the device and system models and synthetic environment representations that the User will likely use in day-to-day applications. Results validation accomplishes two purposes:

[26] See the special topic on Validation for additional information.

- testing the HBR against the acceptability criteria to assure that it performs to the standards required by the purpose
- providing an opportunity to evaluate the interaction of all the component behaviors of an HBR over time against reality

Results validation normally includes separate periods of time to

- conduct acceptability criteria testing
- exercise scenario vignettes to assure that the HBR performs reasonably under operational conditions

One should validate HBR results in the context of the results validation for the entire simulation system. As a result, any special tasks performed to validate HBR components should augment rather than replace normal results validation procedures.

Additional information about results validation can be found in the V&V Agent core documents (i.e., V&V Agent for new, legacy, federation).[27]

## *How important is results validation of HBRs?*

HBR results validation is the last and most important step in system validation. It should expose the final product from the development process to intensive examination under the same conditions expected during its use. HBRs, because of their complexity, present special validation problems that often only testing can solve. However, HBR testing is itself a complex process full of pitfalls that may not exist in simulations of objects for which straightforward physical laws determine their behavior.

## *What makes HBR results validation so difficult?*

The following factors make HBR results validation difficult:

- The space of possible behaviors is typically very large, even for relatively simple HBRs
- The surfaces in that space that constrain possible behavior are generally nonlinear
- The HBR knowledge elements or behavior moderator models may introduce stochastic aspects
- The HBR's behavior can be very sensitive to initial and boundary conditions thereby making that behavior chaotic

[27] See the core documents on V&V Agent for New Simulations, V&V Agent for Legacy Simulations, and V&V Agent for Federations for additional information.

Testing of an HBR drives the simulation over paths through the possible behavior space. A large space means that the number of possible paths is very large thereby making the possibility of visiting all points in that space infeasible to impossible. If the surfaces in that space behaved linearly then only a few parts of the space could be sampled and the rest of the behavior inferred by interpolating between the sampled points. Nonlinear, stochastic and chaotic spaces with unknown dependencies means that sampling and interpolation cannot be used. Thus, theoretically, the only way to completely test an HBR is to experimentally drive over every possible path in its behavior space.

## *How should one test to discover an HBR's validity?*

HBR test plans should strictly reflect and be limited in scope by the User objectives and supporting scenarios. Software engineers should develop, execute, and document test plans per normal software development procedures. However, SME(s) should be included in the testing process by either direct participation or by review of the testing reports.

In addition to testing behaviors for faithfulness to requirements, SMEs should attempt to input unusual parameter values and to create opportunities for the intersection of two or more behaviors to determine if the logic will handle the conflict gracefully as well as doctrinally. One should never assume that the User will always use the simulation in accordance with the assumptions inherent to the design. The SME evaluation should strive to identify as many pitfalls as possible during testing.

The SMEs should conduct separate tests in which the user documentation is stressed and the simulation is challenged with unorthodox uses of the behaviors. The intent is not to stress the software for functionality for which it was not designed, but to seek out unintended, aberrant behavior from unexpected synergistic effects of behaviors or data before they are discovered during the course of the application.

Testing HBRs beyond their limits is easy. The real trick is sufficient testing within the limits of their expected use. Testing should be most detailed where behavior interactions are their most complex. This is where the surfaces of likely behavior are most convoluted so isolated sampling will provide the least accurate measures of overall behavior. Identifying these areas requires the most attention from the developing software engineers. In addition, intense testing should be done near the limits of the behavior mechanisms to identify those limits.

All this said, testing provides the least complete means of validation. Each test only supplies information on the system behavior over a single path or behavior thread. Extrapolating the validity of that information to other untested behavior threads, unless their relationships are formally understood, is both dangerous and unjustified especially for complex areas in the behavior space (i.e., those involving many interactions).

Therefore, especially during the exercise of scenario vignettes, efforts should be made to select situations to challenge the integrity of the behavioral development throughout the scenario. Simply stated, the evaluators should make every attempt to "break" the simulation. This is especially true for interactive simulations where the Users will be interacting with the simulation directly (e.g., human-in-the-loop (HITL) simulations), possibly in the free play mode. Unfortunately, Users always seem to find the set of commands that "break" the simulation. Every attempt should be made to exercise the robustness of the simulation during results validation in an attempt to preempt discovery of fatal command sequences and/or aberrant behaviors during the application.

## *What purposes do SMEs play in HBR results validation?*

As the User's representative, SMEs provide crucial support to HBR results validation. The importance of SME involvement in results validation is second only to the importance of SME support for requirements development. SMEs optimally take active roles in

- developing the HBR test plan
- designing the HBR testing scenarios
- conducting the testing and collection of data
- assessing the test results against the acceptability criteria to determine the HBRs validity

The assistance of SMEs to the software engineers during results validation can save time and money. The SMEs can expose conceptual errors and suggest directions to overcome those errors to the development team.

## *What does behavior scoping involve and what purposes does it serve?*

Behavior scoping identifies a set of behaviors (i.e., those things that the simulated people do) sufficient to achieve only the User's objectives. This reduces the behavior space that must be validated to a fraction of that required for more general situations. If an existing HBR is being used then this set may also be a small fraction of the behavior that simulations can actually generate. This step makes HBR validation manageable and, in many cases, feasible.

An exercise's scenario also contributes to reducing the behaviors that must be validated for a particular purpose. The spanning set of behaviors can be anticipated from the potential or known scenario(s) to be executed to achieve the User's objectives (e.g., there is no need to develop and validate friendly offensive behaviors if the scenarios are intended to exercise only friendly defensive operations). The scenarios create the mission, enemy, troops, terrain and time (METT-T) context that in turn drives the

behavior requirements. The user and problem domain requirements drive the level of resolution of those behaviors.

## *How can problems discovered during HBR testing be resolved?*

The following actions can resolve problems during HBR testing:

- The development team can correct the knowledge base or the behavior engine or both. These areas and all those to which they are coupled should then be re-tested to assure resolution of the problem and that the fix created no new problems.
- Further testing can be conducted to define the limits of the problem and the development team can document the problem area and its limits. The User can then avoid excursions through this part of the behavior space through judicious design of the exercise scenarios.
- The problem behavior can be redefined as a feature of the simulation, one that more accurately represents the idiosyncrasies of human behavior.

Each of these alternatives has benefits and downfalls. Repair or even diagnosis of problems in complex HBRs may be extremely difficult. If this situation imposes unacceptable cost or schedule risk then the other two alternatives may be explored. However, the User should be involved in making these decisions so they clearly understand what HBR capabilities will be delivered to them.

## *How can one focus HBR results validation to improve its cost effectiveness?*

In order to develop a complete picture of any HBR's capabilities and limitations, testing should cover the entire space of possible situations and the system's responses to those situations. This is clearly uneconomical for most HBRs and their applications. The scope of the testing, and ultimately results validation, effort can be limited in a few basic ways thereby improving the cost effectiveness of this effort:

- Use test scenarios derived from the User's objectives to reduce the volume of the behavior space searched thus testing only those capabilities required to meeting the User's objectives (i.e., don't test areas that will not be used)
- Concentrate testing upon the behavior that can most affect the operational results thus testing only those capabilities exercised by the specified scenarios (i.e., don't test capabilities that are not needed)
- Compare observable HBR behavior with experimental data on actual human behavior and performance to determine error/accuracy and for a sanity check
- Concentrating detailed testing in areas of the behavior space that will be visited most frequently (i.e., test what will be seen the most often)

- Concentrating detailed testing in those areas where the most interactions between behaviors occur (i.e., test the most complex parts)
- Carefully testing the non-HBR simulation components to reduce the possibility of their obfuscating the diagnosis of HBR problems

The last two suggestions focus testing in places with the greatest risk - where HBR errors are likely to have the most impact upon the User and where HBR errors are most likely to occur.

Those areas most likely to affect the HBR's operational validity include those areas where

- scenarios will visit the most (i.e., errors will most probably show there)
- multiple behaviors interact (i.e., built upon the assumption that interacting nonlinear behaviors will create even more convoluted nonlinear behavior)
- most dependencies between properties in the behavior space exist
- behavior space property changes are most sensitive to one another
- behavior is nonlinear (i.e., those areas where errors can hide or be misinterpreted)
- component behaviors transition from one to another nonlinearly
- interactions with environment are most complex
- behavior is stochastic (e.g., probabilistic sensing)

These areas are where problems that can affect valid behavior are most likely to occur.

## *What purpose does a behavior auditing capability serve in HBR validation?*

A behavior auditing or tracing capability is crucial to HBR development and validation. Without specific audit information, it is impossible to know for sure what is occurring and why. A correct behavior could occur for the wrong reason. The danger of this occurrence is that improper behavior may occur as a result of the unexpected synergy of intersecting behaviors.

Only behavior auditing can provide the insight necessary to determine if an observed aberrant behavior truly results from a programming error or if it represents acceptably realistic behavior. Such information can clearly influence many HBR validation decisions. An audit capability that generates specific and detailed information on the observed behavior and its causes should be available for proper behavior development and validation.

The audit trail generated by such a capability should describe each decision and all of the conditions that lead to that decision. This enables logic threads to be examined in detail and makes the development team more efficient. Such auditing associated with rule-based or production-reasoning systems is called rule tracing or behavior explanation. The artificial intelligence community has developed very sophisticated behavior explanation technology for a number of different reasoning approaches. The developer should be encouraged to incorporate that technology as an integral part of the HBR if they had not intended to do so.

## *How can validating an HBR's conceptual model help its results validation?*

Validating conceptual models could provide leverage for results if people take the trouble to build them for HBRs. HBRs are often constructed incrementally with no complete conceptual model to start. If a conceptual model exists then validating it can identify areas where one should concentrate testing including those areas where

- scenarios will exercise repeatedly
- behavior complexities exist,
- behavior unknowns exist

## *How can separately validating an HBR's knowledge base help its results validation?*

For HBRs, knowledge bases are just computer programs interpreted by the behavior engine and their validation is a critical element of the HBR validation process. Their validation could improve testing by identifying areas where

- behavior complexities exist
- observable behavior is sensitive to input state
- scenarios repeatedly exercise
- simple behaviors dominate and data interpolation is possible

If the knowledge base has the right form then existing KBS validation tools can help its validation.

## *How do interoperability issues affect the validity of HBR results?*

Simulation infrastructures permit HBRs to interact with their simulated worlds and so achieve technical interoperability but these infrastructures cannot guarantee substantive interoperability [Dahmann, et al., 1999]. Harmon and Youngblood [2001b] introduce several criteria through which to identify substantive interoperability problems that can affect the validity of the integrated simulation. However, HBRs seem to confound using

these interoperability assessment criteria because the interpretation of their fidelity appears unclear.

Harmon and Youngblood [2001a] shows that one can assign fidelity to HBRs either by evaluating them as black boxes or by assessing the fidelity of their internal workings. Both of these paths yield information through which to assess their interoperability with the simulations that create the world with which they interact. In addition, one can treat the problems associated with qualitative property values, common in HBRs, by generalizing the notions of domains and ranges away from continuous metric spaces to sets of acceptable values. Harmon and Youngblood [2001a] discuss the use of these sets to define dependency sensitivities, precisions and errors for properties that rely upon symbolic values to represent their states.

HBRs must interact with the simulated world both by receiving input data from that world and by controlling properties of that world through their action output. These paths create the opportunities for interoperability problems due to functional dependencies. HBRs can interact with strictly physical objects in their simulated worlds or with other HBRs. The primary distinction between these interactions resides in the ability of HBRs to build and manipulate internal representations of the external world state from their input.

One can test for functional dependency generated anomalies with simulations of strictly physical objects by applying the criteria suggested by Harmon and Youngblood [2001a] to the physical object models and to the HBR as a single black box object with many dependencies. This simplifies interoperability testing by ignoring the specifics of the HBR internals. However, black box evaluation may identify the existence of substantive interoperability problems but provide no further insight into the means available to resolve them.

Further, when testing interoperability with other HBRs, one should examine the ways in which their knowledge bases and internal state representations can interact. This requires assessing the fidelity of the interacting HBRs' internal mechanisms. At this level, one applies Harmon and Youngblood's [2001a] anomaly detection criteria to the fidelity of all of the dependencies represented in each interdependent knowledge base element (e.g., production rule).

Evaluation at this level can uncover functional dependency problems between simulations as well as between the elements of the knowledge base. This also provides enough information to detect interoperability anomalies arising from manifold representations and to permit exploring different solution options. These tests, in effect, check the semantic consistency of the interacting HBR knowledge bases. Such semantic inconsistencies can create very subtle and difficult to diagnose interoperability problems. In these cases, one uses the same criteria suggested above but applies them to the abstract dependencies incorporated in the knowledge base to check their semantic consistency.

An HBR's knowledge base and internal state representation essentially reflects that entity's internal simulation of the simulated world with which it interacts. This creates one or more manifold representations between which interoperability problems can arise. Each HBR involved in a system creates these representations, often quite extensive, and the combinatorics can be staggering if interacting HBRs can represent each other's internal state representations (i.e., what I think I know of what you think you know about the world and, perhaps, me).

HBR technology has approached this level of complexity just recently in various simulations that reason about the other actors in their worlds. Ignoring an HBR's internals means overlooking the existence and impact of these extensive and complex manifold representations. Black box derivations of HBR fidelity cannot provide the information needed to assess the impact of these manifold representations upon the substantive interoperability of simulations incorporating HBRs. One can only detect these interoperability problems by applying the anomaly detection criteria at the internal level. Evaluating the fidelity of an HBR's internal operations provides much more information at the price of considerably more effort.

Semantic consistency and manifold representation problems commonly arise in real people and their organizations and can dramatically affect their behaviors. The existence of these interoperability problems in HBRs can actually improve the validity of the entire simulation system. But, one should always be certain that the interoperability anomalies arising from these problems truly represent the situations that actually occur in the real world and are not simply manifestations of substantive interoperability anomalies that occur only in simulation. Black box evaluation may not provide sufficient insight to make this assessment possible.

## *What can one do when validation activities find HBR behavior anomalies?*

One essentially has three options when results validation activities reveal behavioral anomalies. They may modify the

- operational scenarios to avoid those areas where the anomalous behavior occurs (the easiest of options)
- contents of the knowledge base to correct the anomalous behavior (the next easiest option)
- behavior engine to correct the problem (the hardest option)

The extent of the anomalous behavior and the degree to which the Users can tolerate that behavior largely determine the option that best suits the situation. Changes to the knowledge base or behavior engine will probably require additional validation activity whereas changes to the scenarios may not unless the scenario changes drive the HBR into parts of its behavior space that have not been validated for the current purpose.

# References

Adelman, L., Evaluating Decision Support and Expert Systems, John Wiley & Sons, New York, NY, 1991.

Ayel, M.C. & Laurent, J.-P., eds., Validation, Verification and Test of Knowledge-Based Systems, John Wiley & Sons, New York, NY, September 1991.

Bahill, A.T., ed., Verifying and Validating Personal Computer-Based Expert Systems, Prentice Hall, Englewood Cliffs, NJ, 1991.

Dahmann, J. et al., "HLA and Beyond: Interoperability Challenges," Paper No. 99F-SIW-073, Proc. Fall 1999 Simulation Interoperability Workshop, Orlando, FL, September 1999, np.

Geissman, J.R. & Schultz, R.D., "Verification and Validation of Expert Systems," AI Expert, 1988, pp. 26-33.

Grogono, P., Batarekh, A., Preece, A., Shinghal, R. & Suen, C., "Expert System Evaluation Techniques: A Selected Bibliography," Expert Systems, 9, 1992, pp. 227-239.

Grogono, P., Preece, A.D., Shinghal, R. & Suen, C.Y., "A Survey of Evaluation Techniques Used for Expert Systems in Telecommunications," Expert Systems with Applications, 5, 1992, pp. 395-401.

Gupta, U.G., ed., Validating and Verifying Knowledge-Based Systems, IEEE Computer Society Press, Los Alamitos, CA, 1992.

Hamilton, D. & Kelly, K., "State-of-the-Practice in Knowledge-Based System Verification and Validation," Expert Systems with Applications, 3, 1991, pp. 403-410.

Harmon, S. Y., Bibliography of Verification, Validation, Evaluation and Testing of Knowledge-Based Systems, Defense Modeling and Simulation Office, Alexandria, VA, 1998.

Harmon, S.Y., "Validation of Human Behavior Representations," Paper # 99S-SIW-048, Proc. Spring 1999 Simulation Interoperability Workshop, Orlando, FL, March 1999, np.

Harmon, S.Y. & Youngblood, S.M., "Fidelity and Human Behavior Representations," Paper # 01S-SIW-113, Spring 2001 Simulation Interoperability Workshop, Orlando, FL, March 2001, np.

Harmon, S.Y. & Youngblood, S.M., "Leveraging Fidelity to Achieve Substantive Interoperability," Paper # 01S-SIW-114, Spring 2001 Simulation Interoperability Workshop, Orlando, FL, March 2001, np.

Lopez, B., Meseguer, P. & Plaza, E., "Knowledge-Based Systems Validation: A State of the Art," AI Communications, 3(2), 1990, pp. 58-72.

Lydiard, T.J., “Overview of Current Practice and Research Initiatives for the Verification and Validation of Knowledge-Based Systems,” The Knowledge Engineering Review, 7(2), June 1992, pp. 101-113.

Mengshoel, O.J. & Delab, S., “Knowledge Validation: Principles and Practice,” IEEE Expert, 8(3), June 1993, pp. 62-68.

Nazareth, D.L. & Kennedy, M.H., "Knowledge-Based System Verification, Validation and Testing: The Evolution of a Discipline,” Int. Jour. of Expert Systems, 6(2), 1993, pp. 143-162.

O'Keefe, R.M. & O'Leary, D.E., “Expert System Verification and Validation: A Survey and Tutorial,” Artificial Intelligence Review, 7, 1993, pp. 3-42.

Pew, R.W. & Mavor, A.S., eds., Modeling Human and Organizational Behavior, National Academy Press, Washington, DC, 1998.

Preece, A., Shinghal, R. & Batarekh, A., "Principles and Practice in Verifying Rule-Based Systems,” Knowledge Engineering Review, 7(2), June 1992, pp. 115-141.

Preece, A.D. & Suen, C.Y., eds., “Special Issue on Verification and Validation,” Int. Jour. of Expert Systems - Research and Applications, 6(2/3), 1993.

Preece, A.D., “Validation of Knowledge-Based Systems: Current Trends and Issues,” Knowledge Engineering Review, 10(1), 1995, pp. 69-71.

Smith, S. & Kandel, A., Verification and Validation of Rule-Based Expert Systems, CRC Press, Boca Raton, FL, nd.

Veit, C.T. & Callero, M., Criteria for Validating Human Judgments and Developing Behavioral Representation Models, The Rand Corporation, Santa Monica, CA, 1993.

Zlatareva, N. & Preece, A., “State of the Art in Automated Validation of Knowledge-Based Systems,” Int. Jour. of Expert Systems – Research and Applications, 7(2), 1994, pp. 151-167.

## *RPG References in This Document*

select menu*: Core Documents*, select item: “V&V Agent Role in the VV&A of Federations”

select menu*: Core Documents*, select item: “V&V Agent Role in the VV&A of Legacy Simulations”

select menu*: Core Documents*, select item: “V&V Agent Role in the VV&A of New Simulations”

select menu*: Reference Documents*, select item: “Human Behavior Representation Literature Review”

select menu*: Reference Documents*, select item: “V&V Techniques”

select menu*:* *Special Topics*, select item: "Conceptual Model Development and Validation"

select menu*:* *Special Topics*, select item: "Paradigms for M&S Development"

select menu*:* *Special Topics*, select item: "Requirements"

select menu*:* *Special Topics*, select item: "Subject Matter Experts and VV&A"

select menu*:* *Special Topics*, select item: "Validation"

§ § § § § § §