Human Behavior Representation (HBR) Literature Review

RPG Reference Document

8/15/01[1]

Table of Contents

Abstract	i
Introduction	1
VVE&T of KBS Technology	1
Theory	2
Techniques and Tools	2
Problems	3
Technology	3
Validation of HBR Technology	3
Domain Correspondence	5
Psychological Correspondence	5
Physiological Correspondence	6
Increasing HBR Validation Capabilities	7
Apply KBS VVE&T Technology More Widely to HBRs	7
Broaden Correspondence Testing of HBRs	8
Develop Accessible Databases of Human Performance Data	9
Conclusion	10
References	10
Survey Articles	11
RPG References in This Document	11

[1] This document replaces the 11/30/00 version. It contains minor editorial and formatting changes.

This document corresponds to the web version of the VV&A RPG Reference Document of the same name and date. It has been modified to make it suitable for printing.

ABSTRACT

The inherent nonlinearity and enormity of possible human behavior responses presents extremely complex challenges for the validation of human behavior representations (HBRs).[1] Existing technology for validating knowledge-based systems provides some theory, techniques, tools, and experience to address these challenges. However, these resources apply primarily to rule-based expert systems. Current HBR validation techniques for military models and simulations rely upon the reviews of knowledge bases and observable behavior by subject matter experts (SMEs)[2]. However, most SMEs can only validate knowledge representations with such "readable" formats as production rules or when interpreted by software developers. HBRs with more abstract knowledge representations can only be validated by examining the observable behavior generated when executing an application's scenarios. Then, an SME can only completely validate the observable behavior by exhaustively exploring the entire space of possible responses within the context of a given scenario, an impossible task for even modest human representations. These serious deficiencies in existing HBR validation technology call for new approaches. This document suggests that accurate HBRs will correspond to human behavior at five levels of representation: domain, physiological, psychological, organizational and physical. An HBR with correspondence at all five levels is often better than an HBR with correspondence in fewer. This document defines each level of correspondence and compares them to existing HBR capabilities. Results show that no HBR currently exists with complete correspondence at all five levels.

[1] See the special topic on Validation of Human Behavior Representations for additional information.
[2] See the special topic on Subject Matter Experts and VV&A for additional information.

Introduction

As information system technology advances, constructing fine-grained simulations of human behavior for a variety of situations becomes more feasible and practical. Developers have built many simulations of human behavior, primarily cognitive, and the sophistication of these systems continues to improve. However, validation of human behavior representations (HBRs)[1] has always been difficult. Human behavior manifests a highly complex fabric of effects coupled over many orders of magnitude, a property shared by chaotic systems. Small situation changes often create wildly different responses in the same system. Thus, validation of HBRs, even for simple tasks, can be extremely difficult because of the numerous behavioral paths that need to be explored for any given application. The lack of well-established techniques and tools to support HBR validation further complicates the situation.

This document examines the technology available for validating HBRs, identifies current deficiencies, and recommends approaches to overcoming these deficiencies. The technology for validating HBRs include the current capabilities for verifying, validating, evaluating and testing (VVE&T) knowledge-based systems (KBSs) and the experience in validating existing HBRs for a variety of applications.

VVE&T of KBS Technology

A survey of the current literature about VVE&T of KBSs was conducted to assess the state of the art in this area and to determine its applicability to the validation of HBRs. Several survey articles, listed in the reference section, and books [Ayel and Laurent, 1991; Bahill, 1991; Gupta, 1992] provide an overview of this technology area. The table below presents a summary of the results to date of this literature survey.

Summary of Results from KBS VVE&T Literature Survey	
Analysis Property	**Property Value**
• Number of references found	531
• Years covered	1971-1998 (27 years)
• Number of workshops found	19 (1988-1997)
• Number of books found	4
• Number of different authors found	684

The amount of existing literature about artificial intelligence (AI) and the length of time it has been studied were surprising. The AI community has been interested in validation

[1] See the special topic on Validation of Human Behavior Representations for additional information.

for a long time. However, the number of relevant references written during this period is only one measure of the state of the art.

The amount of theory, the number of different techniques, the number of available tools and the amount of experience better represent the state of the art in any technical area. The table below summarizes this information as drawn from the literature surveyed for the VVE&T of KBSs.

State of the Art for KBS VVE&T from Existing Data		
Field Characteristic	**Number Found**	**Specific Categories Studied**
Theory	16 references	• data selection, verification, validation, testing
Techniques	41 techniques	• logic; optimization; classification; transformation; graph theory; empirical, heuristic, and formal methods; modeling and simulation
Tools	60 tools	• specification, verification, validation, refinement, testing, performance evaluation
Problems	25 problems	• integration, knowledge conditions, specific representations, specific architectures, V&V processes
Technology Experience	115 references	• medical, financial, analytical chemistry, management decision aiding, space, telecommunications, computer design, laboratory data analysis, manufacturing, scheduling, mineral exploration

Theory

While the theory relevant to the VVE&T of KBSs seems comparatively small (only 16 references), it addresses all of the important problems. In addition, developing theory to underlie the behavior of KBSs is a challenging task that has only recently seen some promising advances. Until a comprehensive and consistent KBS theory exists, theory supporting VVE&T of KBSs is likely to remain as loosely coupled concepts. However, a wide variety of verification and validation (V&V) techniques and tools have been proposed and tested.

Techniques and Tools

The literature surveyed discusses 41 different techniques that can be grouped into the nine categories shown. The survey uncovered 60 tools that support different aspects of VVE&T of KBSs. These varied 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. The literature surveyed also revealed an enormous amount of experience in VVE&T of KBSs with 115 different references for diverse applications. By far, most of

this experience was related to medical applications where the results from any KBS can have life-threatening consequences.

Problems

The surveyed tools and techniques, shown in the table above, addressed 25 different problems associated with KBSs. These were grouped into five broad categories. The table below shows the specific problems associated with each category.

Summary of KBS V&V Problems Studied	
Problem Category	**Specific Problems Addressed**
Knowledge Base Integration	• completeness/coverage, consistency/coherence, redundancy
Knowledge Conditions	• incomplete, multi-level, modular, uncertain, incorrect
Specific Representations	• nonmonotonic, case-based, tabular, equations, weighted rules, control/meta-knowledge, dynamic properties
Specific Architectures	• blackboard systems, expert system shells, multi-agent systems
V&V Processes	• automatic refinement, knowledge base verification, subjective criteria, large knowledge bases, wide domains

Technology

KBS VVE&T technology is an important resource for the validation of HBRs. This technology is relatively well developed and broad, addressing a number of different fields and areas of experience. Many of the tools discovered could be applied to future HBRs with appropriately selected implementation strategies. However, of all of the references found, only one was directly related to evaluation of HBRs [Veit and Callero, 1993]. The largest amount of the work referenced related explicitly to expert and decision support systems.

Most of the technology reviewed in this survey specifically addressed KBSs using production rule knowledge representations. No other knowledge representations were identified. This current focus on rule representations has limited impact on the validation of HBRs because production rules are the most common representations they use. However, because most existing HBRs were not designed to take advantage of any of the existing KBS VVE&T tools, it would be necessary to modify either the HBRs or the tools to directly benefit from the 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 (e.g., stress, injury, emotion).[2]

[2] See the special topic on Validation of Human Behavior Representations for additional information.

Validation of HBR Technology

Several HBRs have been developed for a variety of purposes. A recent National Research Council study, *Modeling Human and Organization Behavior* [Pew and Mavor, 1998], provides an excellent survey of the validation of many of the existing and developing HBRs. The table below summarizes and compares the different HBR validation approaches discussed in the Pew and Mavor study.

Three different categories of HBR validation: domain correspondence, psychological correspondence and physiological correspondence are shown in the table below and described in the sections following.

Comparison of the Validation of Different HBRs

System	Domain Types	Correspondences			Validating Data Sources
		Domain	Psychological	Physiological	
ACT-R	submarine TAO & Aegis radar operators	X	X		• human behavior data
COGNET	anti-submarine warfare	X			• human behavior data
EPIC	computer interaction tasks	X	X		• human behavior data
HOS			X		• validated theory
Micro SAINT	helicopter crew, ground vehicle crews, C2 message, tank maintenance & harbor entry operations	X			• human behavior data
MIDAS	757 flight crew	X			• human behavior data
Neural Networks			X	X	• validated theory • human behavior data
OMAR			X		• validated theory • human interaction
SAMPLE			X		• validated theory
Soar	air traffic control, test director, automobile driver, job shop scheduling	X	X		• validated theory • human interaction • human behavior data
ModSAF	ground warfare	X			• human interaction
CCTT SAF	ground warfare	X			• human interaction
MCSF	small unit operations	X			• human behavior data • human interaction

Comparison of the Validation of Different HBRs					
System	**Domain Types**	**Correspondences**			**Validating Data Sources**
		Domain	**Psychological**	**Physiological**	
SUTT CCH	small unit operations	X	X		• human behavior data • human interaction
IFOR (see Soar)	fixed & rotary wing air operations	X	X		• validated theory • human interaction • human behavior data

Domain Correspondence

Domain experts know the results that typical human behavior in their particular domains should produce. This knowledge permits those experts to examine HBR knowledge bases and observe HBR performance in their domains and assess, often qualitatively, how realistically the necessary human behaviors are represented. Developers and Users of HBRs have applied domain expert review (i.e., domain validation) more often than any other validation technique.

Domain correspondence testing was first widely employed in expert system construction (i.e., empirical techniques). This form of validation is equivalent to asking if the HBR produces results indistinguishable from those of human experts and corresponds to what some have called the Turing test (although most of the recent Turing test interpretations have deviated significantly from what Alan Turing originally intended as a test of machine intelligence).[3] The assessment of HBR performance through interactions with experts is denoted in the Comparison of the Validation of Different HBRs table as human interaction data sources. The qualitative nature of this correspondence test, together with the need to explore a large portion of the problem space (because of the inherent nonlinearities of human performance), make this the weakest form of HBR validation.[4]

Some quantitative experimental data exist on actual human performance in various battlefield situations (e.g., data from instrumented ranges) and on humans performing very specific cognitive tasks. The results produced by HBRs can be compared to these data rather than the opinions of experts. The Comparison of the Validation of Different HBRs table denotes this approach to domain correspondence testing as human behavior data sources. This performance data, when available, strengthens the results of domain validation. Regrettably, domain-specific data are often very sparse and apply to narrow situations. Experimental conditions for data collection are often poorly controlled and characterized. Such problems tend to weaken the validation effort done against the source data.

[3] See the reference document on V&V Techniques for additional information.
[4] See the special topic on Validation of Human Behavior Representations for additional information.

Regardless of whether domain validation is quantitative or qualitative, its specificity requires testing over much of the problem space to assure acceptably correct performance in loosely controlled simulation scenarios (e.g., free play). This extent of testing can be impossible or, at the very least, uneconomical for highly capable HBRs. Because of the delicate balance between validation practicality and risk, anomalous behavior could occur (and often has) when least desired. This fact indicates the need to supplement domain validation with other information.

Psychological Correspondence

There is a vast body of knowledge about the psychology of humans that includes numerous abstract models of many different aspects of human behavior. There is also an enormous volume of published experimental data on actual human performance under various circumstances, which validate these models to some degree. This knowledge enables the testing of psychological 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 baseline performance against which to compare HBR behavior. Experimental data can completely establish a baseline or can augment a baseline created by psychological models. Then, a set of carefully controlled experiments produces HBR performance data. Comparing these data against the baseline validates the models underlying the HBR. Psychology professionals can be employed to qualitatively validate complex behavior in the same way that domain experts are used to evaluate domain correspondence. These professionals effectively determine if the HBR performs in a psychologically realistic manner.

Psychological correspondence testing creates stronger validation than domain validation because of its linkage to the underlying models of psychological phenomena. Thus, the entire problem space need not be explored as in testing domain correspondence. In addition, the experimental data from which baselines are drawn have more likely been obtained under carefully controlled experimental conditions and are therefore more repeatable than that obtained in domain-specific experiments. This form of correspondence testing might be better suited for integrated models of human behavior that incorporate such effects as stress and emotion. In fact, as shown in the Comparison of the Validation of Different HBRs table, integrated models of human behavior are most likely to be validated through psychological correspondence testing.

Physiological Correspondence

A considerable collection of experimental data and a recently developed verifiable theory of neurophysiological processes have established additional baselines against which to compare HBR performance. HBRs have one significant advantage over the actual physiological systems from which these data originate: their detailed workings

are easier to directly observe. Simulations that have physiological correspondence are more likely to behave like real people especially under conditions where non-neurological physiology contributes (e.g., fatigue and injury).

Testing physiological correspondence equates to asking a neurologist, neurosurgeon, or physiological psychologist to evaluate a system's performance by looking at the behavior of the human components it represents. This sort of evaluation is much closer to what has traditionally been done to validate physical system representations (i.e., nonhuman systems). In the past, this kind of validation was 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). Primarily neural network approaches to HBR have been validated through physiological correspondence testing and these validation efforts have been limited to the relatively constrained performance of nonspecific neurons.

All of the validation techniques applied to existing HBRs have significant limitations:

- 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 the validation of HBRs, especially for simulation applications. 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.[5]

Increasing HBR Validation Capabilities

Despite considerable relevant technology in the VVE&T of KBSs and the validation of existing HBRs, HBR validation has not yet reached the level of assurance that models of physical systems can achieve. Considering the fragmented technology picture presented here, current users of HBRs can hardly be faulted for performing limited validation of their systems. Three major steps are recommended to change this situation thereby increasing HBR validation capabilities and improving HBR credibility.

- Apply KBS VVE&T Technology More Widely to HBRs
- Broaden Correspondence Testing of HBRs
- Develop Accessible Databases of Human Performance Data

[5] See the special topic on Validation of Human Behavior Representations for additional information.

Apply KBS VVE&T Technology More Widely to HBRs

The artificial intelligence community has already invested considerable resources developing the technology for VVE&T of KBSs. This technology base has significant application potential to HBRs. This application potential can be realized in two ways:

- explore the feasibility and practicality of adapting existing KBS VVE&T theory, tools, and techniques for HBR validation
- promote the development of new HBRs that take advantage of the existing techniques and tools for VVE&T of KBSs

This strategy improves the likelihood of KBS VVE&T technology employment from two ends. Adapting existing VVE&T techniques increases the opportunities for their employment by existing HBRs and new HBRs built upon existing systems. Developing new HBRs that use existing KBS VVE&T technology emphasizes the importance of designing for validation. These two paths will cause the KBS VVE&T and HBR technical communities to grow together so that each can leverage the other's investments and capabilities.

Broaden Correspondence Testing of HBRs

Validation of HBRs through correspondence testing can be expanded in two ways:

- broaden the testing of HBRs for psychological and physiological correspondences
- develop new techniques to test HBRs for sociological and physical correspondences

Psychological and Physiological Correspondence Testing

In general, validation of HBRs by testing their correspondence with psychological and physiological models has been limited. Significant opportunities are available for reducing the time and cost required to validate HBRs through psychological and physiological correspondence testing. For example, there are numerous standardized psychological tests that could be administered to sophisticated HBRs and their results could be compared with statistics obtained from the huge reservoir of human testing results. The techniques for designing psychological correspondence experiments are well understood and well documented by the psychological community. However, some methods need to be developed so the suitability of the various psychological models and experimental data to different HBR applications can be determined.

Recent advances in noninvasive measurement techniques (e.g., MRI, PET) have increased understanding of the linkage between cognitive behavior and physiological

observations and have created a large repository of potential validation data.[6] As this area of experiment improves, comparing these experimental results with HBR designs and performance should become easier and more meaningful.

Physical and Sociological Correspondence Testing

Two additional areas of correspondence that have not yet been widely employed in the validation of HBRs are physical and sociological.

Physical Correspondence

The human neural system consists of a collection of interacting computational devices. Basic physical laws limit the performance of these devices as they do in silicon-based computation. Consequently, the computational performance of an HBR can be compared to the limits predicted by the laws of physics governing the brain's computational activities. Representations that exceed these limits predict the behavior manifested by the brain inaccurately.

Many aspects of observable human performance may originate from these reasonably simple physical limitations. Unfortunately, the physics of computation remains a relatively unexplored area and so lacks the theory and experimental results to create reliable baselines. Validation of the physical limitations represented by a model or simulation has been done extensively for nonhuman systems but never applied, as yet, to HBRs. Nevertheless, humans are physical entities constrained by physical laws and so there must be correspondence between HBRs and the underlying laws of physics for them to produce accurate results. Realizing that humans are subject to the same limitations imposed upon all computational systems can significantly reduce the inherent complexity of validating HBRs at a physical level. Research is still needed to make this level of validation useful. However, physical correspondence provides one more test to help guarantee the accuracy of HBR performance.

Sociological Correspondence

HBRs that replicate groups of interacting people, including disordered groups such as crowds as well as groups operating within some organization structure, must possess sociological correspondence. As with psychological validation, there is a rich body of sociological knowledge from which baselines 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 validation tests. Sociological correspondence testing is similar to psychological correspondence testing in that it can be tested against the observations of psychology and sociology professionals, such as experts in organizational structure and dynamics. Current sociological knowledge permits the testing of group behaviors as well as the interaction dynamics between group members. Testing sociological validity is particularly important

[6] See the reference document on M&S Data Concepts and Terms for additional information.

with simulations of human groups cooperating to perform some task and may not be necessary when representing the actions of a single individual.

Develop Accessible Databases of Human Performance Data

By far the most important deficiency in the current ability to validate HBRs is the lack of an accessible repository of consistent information about human performance in a variety of circumstances. Certainly, much of this information exists and more is being collected all the time. However, the burden of locating appropriate validation data and then transforming that data set into a consistent and meaningful collection rests with the individual validation agent or developer.[7] Having an accessible repository of such data would surely reduce the burden of HBR validation and thereby increase the likelihood of any validation activity. As more HBRs are validated, their results can be added to this repository to further reduce validation costs. In addition, a consistent data collection will encourage the performance of different HBRs to be compared on equal footings. This will provide designers with critical information to assist in making their HBR design decisions.

Conclusion

Resources for HBR validation exist in the technology supporting VVE&T of KBSs and in current experience in validating the few existing HBRs. However, these resources provide, at best, a disjoint set that leaves most developers and validation agents with little choice other than using subject matter experts to search entire problem domain spaces. The inherent complexity and nonlinearity of most HBRs make this choice the most expensive and least satisfactory of any.

In all, five levels of correspondence establish the validity of HBRs. An HBR that has correspondence at all five levels best approximates human behavior for all applications. Most applications 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.[8] The theoretical models and experimental data associated with the psychological, sociological, physiological 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 would not need extensive revalidation for each new application. This would drastically reduce the cost of validating existing HBRs and hasten their widespread employment. Applying existing techniques for VVE&T of KBSs will further improve the validation of HBRs by supplying a crucial body of established theory, techniques, tools,

[7] See the special topics on Data V&V for New Simulations and Data V&V for Legacy Simulations for additional information.
[8] See the special topic on Validation of Human Behavior Representations for additional information.

and experience to HBR developers and validation agents. When brought together, these resources will dramatically improve the quality and reduce the cost of HBR validation.

References

Ayel, M.C. and 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.

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

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

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

Survey Articles

Grogono, P., Batarekh, A., Preece, A., Shinghal, R. and Suen, C., “Expert System Evaluation Techniques: A Selected Bibliography,” *Expert Systems*, 9(4), 1992, pp. 227-239.

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

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

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

O’Keefe, R.M. and O'Leary, D.E., “Expert System Verification and Validation: A Survey and Tutorial,” *Artificial Intelligence Review*, 7, 1993, pp3-42.

Preece, A.D. and Suen, C.Y., eds., “Special Issue on Verification and Validation,” *IJES -- 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.

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

RPG References in This Document

select menu*: RPG Reference Documents*, select item: "M&S Data Concepts and Terms"

select menu*: RPG Reference Documents*, select item: "V&V Techniques"

select menu*: RPG Special Topics*, select item: "Data V&V for Legacy Simulations"

select menu*: RPG Special Topics*, select item: "Data V&V for New Simulations"

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

select menu*: RPG Special Topics*, select item: "Validation of Human Behavior Representations"

§ § § § § § §