SANDIA REPORT
SAND2008-6453
Unlimited Release
Printed September 2008

R&D for Computational Cognitive and Social Models: Foundations for Model Evaluation through Verification and Validation (Final LDRD Report)

Laura A. McNamara, Timothy G. Trucano, George A. Backus, Scott A. Mitchell, Alexander Slepoy

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone:	(865) 576-8401
Facsimile:	(865) 576-5728
E-Mail:	reports@adonis.osti.gov
Online ordering:	http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone:	(800) 553-6847
Facsimile:	(703) 605-6900
E-Mail:	orders@ntis.fedworld.gov
Online order:	http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2008-6453
Unlimited Release
Printed September 2008

R&D for Computational Cognitive and Social Models: Foundations for Model Evaluation through Verification and Validation (Final LDRD Report)

Laura A. McNamara, Exploratory Simulation Technologies
Timothy G. Trucano, Optimization and Uncertainty Quantification
George A. Backus, Exploratory Simulation Technologies
Scott A. Mitchell, Computer Science & Informatics
Alexander Slepoy, Multiscale Dynamic Materials Modeling

Sandia National Laboratories
PO Box 5800
Albuquerque, NM 87185-0370

Abstract

Sandia National Laboratories is investing in projects that aim to develop computational modeling and simulation applications that explore human cognitive and social phenomena. While some of these modeling and simulation projects are explicitly research oriented, others are intended to support or provide insight for people involved in high consequence decision-making. This raises the issue of how to evaluate computational modeling and simulation applications in both research and applied settings where human behavior is the focus of the model: when is a simulation "good enough" for the goals its designers want to achieve?

In this report, we discuss two years' worth of review and assessment of the ASC program's approach to computational model verification and validation, uncertainty quantification, and decision making. We present a framework that extends the principles of the ASC approach into the area of computational social and cognitive modeling and simulation. In doing so, we argue that the potential for evaluation is a function of how the modeling and simulation software will be used in a particular setting. In making this argument, we move from strict, engineering and physics oriented approaches to V&V to a broader project of *model evaluation*, which asserts that the systematic, rigorous, and transparent accumulation of evidence about a model's performance under conditions of uncertainty is a reasonable and necessary goal for model evaluation, regardless of discipline. How to achieve the accumulation of evidence in areas outside physics and engineering is a significant research challenge, but one that requires addressing as modeling and simulation tools move out of research laboratories and into the hands of decision makers. This report provides an assessment of our thinking on ASC Verification and Validation, and argues for further extending V&V research in the physical and engineering sciences toward a broader program of model evaluation in situations of high consequence decision-making.

Table of Contents

INTRODUCTION

In the United States, across government and industry, mathematical modeling and computational simulation have long been perceived as critical to support decision making (Enthoven and Smith 1971). Throughout the postwar era, modeling and simulation played key methodological roles in a range of technology and policy areas, from nuclear weapons science to weather forecasting to defense acquisition. Perhaps it is not so surprising, then, that since 9/11, federal agencies have invested relatively heavily in computational modeling and simulation software to help analysts and decision makers better understand "soft" problems, like insurgency and terrorism. As a 2006 article in IEEE Spectrum pointed out, the United States' national security enterprise seems to be betting that "computers equipped with the right software can give vital insights into the minds and motives of terrorists and the structure and critical links in their organizations" (Goldstein 2006). Such excitement in the revolutionary potential of computing is hardly new; fifteen years ago, Steven Bankes argued that "We are in a golden age of computer modeling for policy analysis," pointing to growth in computational power, creativity in model-builders' research efforts, and a increasing interest in computer-driven analysis among decision makers in multiple fields and agencies (Bankes 1993).

Despite current investment in computational modeling and simulation for policy analysis, there has been less of a consistent effort to develop methodologies and frameworks for systematically evaluating the fidelity and relevance of those models for the problem spaces they are intended to address. Addressing the reasons for this absence of systematic investment in modeling and simulation software evaluation is beyond the scope of this report, but we do see this as a major issue for policymakers who are looking to models as inputs for decisions that may have consequence for human lives. In such fields as computational science and engineering, operations research, and avionics, modeling and simulation ***evaluation*** **–** also commonly referred to as "***Verification and Validation***" **–** is a mature field of research and development. It has long been recognized in computational science and engineering that "[computer] programs whose malfunction would have severe consequences justify greater effort in their validation" (Adrion, et al. 1982).

Accordingly, the goal of the CS&T V&V LDRD is to explore the **possibility of mapping the verification and validation methodologies developed under the Advanced Simulation and Computing (ASC) program to computational simulations of human cognition and/or behavior**. More specifically, we seek to extend methodological thinking and processes for evaluating computational science and engineering models into the realm of computational simulations of social and/or psychological systems. The specific case that our team has engaged is the Memory and Reasoning (M&R) modeling LDRD project being pursued in 6341, under the leadership of Michael Bernard. The goal of the M&R LDRD is to create a computational simulation of "…how the hippocampal system might process information acquired during learning experiences leading to the consolidation of declarative memories" (Bernard 2008). The design document describing the M&R project and broad structure of the intended computational model is attached as an appendix to this report. This document is a starting point for making the broad thinking presented in this report about the proper context and conduct of computational model evaluation specific to the needs of the M&R project.

We saw, and continue to see, the M&R project as an opportunity to partner with an interdisciplinary group of social and computational scientists, to learn about the process they are engaging in for their modeling work, and to identify ways that verification and validation methodologies developed in the computational science and engineering realm might be extended to other disciplines. This report summarizes our findings.

VERIFICATION AND VALIDATION: A BRIEF REVIEW OF THE INTERDISCIPLINARY LITERATURE

Verification and Validation (V&V) refer to a set of interrelated, mutually reinforcing evaluation methodologies that address the performance of computational simulation software in modeling a real-world phenomenon of interest. Generally speaking, verification evaluates the internal correctness of the software, while validation assesses the correspondence between the software and the real-world phenomenon it is addressing. For example, from a strictly software-oriented perspective, the two terms are usually defined in terns of requirements:

- *Verification* – demonstration that requirements (for software) are correctly implemented.
- *Validation* – demonstration that requirements (for software) are correct.

As we shall discuss below, achieving these "demonstrations" is still both a matter of degree, relative to the intended application; as a general rule, the higher the degree of risk and consequence in the application area, the more effort goes into the process of defining, accumulating, and using evidence. This specification is a little more general than the V&V definitions used in other realms: e.g., computational science and engineering, for which verification and validation are activities associated with partial differential equations and their solutions. In this realm, which "solving equations correctly" and "solving the correct equations" can be cast strictly as requirements implementation and correctness issues.

As we discuss below, verification and validation are research topics in many disciplines where computational methodologies are being or have been adopted to support knowledge production, including the social and the physical sciences. For example, in the computational social science literature, authors provide examples of specific techniques for model evaluation: *docking* (Axtell, et al. 1996; Carley 1996), *calibration* (Fagiolo, et al. 2006), *face validation* (Carley 1996; Fagiolo, et al. 2006). However, the most systematic methodological research in verification and validation has taken place in disciplines where modeling and simulation software has been developed to automate and/or support human decision-making – for example, in avionics, artificial intelligence, nuclear power, and nuclear weapons certification.

In this section, we very briefly review the voluminous literature on modeling and simulation evaluation methodologies from several disciplines. We stress that this is a brief review of the literature we located, included as a framing discussion prior to setting out our argument for the interdisciplinary applicability of V&V frameworks from computational science and engineering to other modeling and simulation application areas. Readers interested in exploring the literature on verification and validation can turn to Appendix A, which contains an extended bibliographic list of key references related to the verification and validation of computational modeling and simulation software.

What is a Model? What is a Simulation?

Before embarking on a discussion of verification and validation, we believe it is necessary to provide some clarification about the words *code*, *model*, and *simulation*. These words are often used interchangeably, leading to term confusion. Osman Balci differentiates between a model as "…a representation or abstraction of something such as an entity, a system, or an idea," while a simulation is "the act of…exercising a model under diverse objectives…" (Balci 2003). Similarly, in this paper, we use the word *model* to designate a representation or abstraction that captures key characteristics of a phenomenon of interest. Models are not necessarily computational; they may be physical, narrative, mathematical, logical, or graphical. In this paper, for example, we will refer frequently to descriptive (or foundational) *conceptual models* as a starting point for verification and validation (for a discussion the relationship between conceptual or mental models and computational models, see (Forrester 1971; Forrester and Senge 1980).

Simulations do not necessarily involve computers, either. A simulation may involve real-world processes; for example, a fire drill simulates the process of vacating a building during an emergency. However, when computers are involved, simulations typically involve the exercise of mathematical models, if only because computers are uniquely able to combine data (inputs) with mathematical representations of real-world phenomena (models) in "computer time" (time steps) to generate data (output).

In developing a computational model and simulation, the researcher identifies and specifies key attributes, or *parameters*, that best describe the phenomenon under study. The researcher then instantiates these parameters and the relationships among them in *code* or *software:* this is the ordered set of instructions through which a logical or mathematical model is rendered computationally tractable. Once the model is instantiated in code, the researcher uses the code to run a *simulation:* an *in-silico*, dynamic representation of some aspect of reality, which enables the researcher to learn more about the real-world behavior of a phenomenon of interest.

Models, codes, and simulations all aim at creating interactive artifacts that amplify human cognition through the quantitative and qualitative, for example visual, representation of information about system change and transition, enabling discovery, explanation, even prediction (Card, et al. 1999). Accordingly, in this paper, when we speak of modeling and simulation tools (or modeling and simulation applications, code or software), we are speaking of models that are specifically designed to be exercised dynamically as simulations, often to enable the exploration of a problem space that might not otherwise be experimentally accessible. This latter objective is emphatic in human interaction research.

Verification and Validation in Computational Social Science

The epistemic culture of computational social science is not one that routinely engages in rigorous or systematic evaluation of the models it produces. Axtell et al, for example, bemoan the absence of evidence that would help define a clear "domain of validity" for social science theories developed using computational models, analogous to that which exists for more greatly mathematized theories. Lacking comparative methodologies, they argue, it is difficult for computational social scientists to assess when one model fits a dataset better than another model; or if one model is actually a special case of another model. Models in the social sciences tend to be "constructed entirely de novo," which they attribute to the fact that "computational modeling offers a striking opportunity to fashion miniature worlds, and this appeals to powerful creative impulses within all of us." In the absence of structured, comparative reference to models that already exist, this highly individualized "creative impulse" makes the cumulative construction of knowledge difficult (Axtell, et al. 1996).

That said, within the computational social science community, there seems to be an increasing recognition that model evaluation represents an understudied problem. Since the mid-1990s, researchers in economics, sociology and political science who use computational techniques have begun pursuing methodological research in the evaluation of agent-based and social network models (Axelrod 2003; Axtell, et al. 1996; Carley 1996; Fagiolo, et al. 2007; Fagiolo, et al. ; Fagiolo, et al. 2006; Polhill, et al. 2005; Wilenski and Rand 2007; Windrum, et al. 2007). However, much of the work on verification and validation in this community focuses in individual methods for either verification or validation, and uses case studies to demonstrate the applicability of one method or another for some aspect of model evaluation. In other words, there is little discussion about developing and promulgating formal standards for addressing the problem of model evaluation. Nor is there much visible interest in formalizing systematic and regular approaches to model evaluation, largely because – as opposed to computational engineering – computational social science tends to be pursued by individual or small groups of researchers as an area of methodological activity, primarily in the academic community.

This means that comprehensive and consistent frameworks for developing, comparing, and evaluating models of social phenomena are rare. The computational social science community is recognizing this problem as well: for example, the *Journal of Artificial Societies and Social Simulation* recently published an article calling for the establishment of a collaboratively created, freely available, community-owned modeling framework, analogous to that developed by climate scientists. The authors argued that the absence of a common framework in the agent modeling community makes it difficult to perform model comparisons and conduct rigorous evaluation. A community modeling effort could provide a standard platform for building, disseminating, and evaluating agent-based models, and for promoting their use in a broader swath of the social science community (Na Alessa, et al. 2006).

When it comes to verification and validation, computational social science is arguably sitting in something of a disciplinary bubble. This will sound strange to many of our colleagues who see social modeling as an example *par excellence* of interdisciplinary research and development (Agar, et al. 2004). Certainly computational social modeling brings social scientists into interesting partnerships with computer scientists and mathematicians. However, rarely do practitioners in computational social science engage systematically with the six decades' worth of research that physicists, weather forecasters, climatologists, and engineers bring to the problem of developing, using, and evaluating computational simulations (e.g., (Oberkampf, et al. 2004). Moreover, there is research in verification and validation in cognitive science – i.e., artificial intelligence and neural network modeling – that could also be leveraged in important ways to address the evaluation of computational social modeling and simulation efforts. We believe that a purposeful intersection between computational social science and these more established application areas will do much to reveal methodological and epistemological challenges facing the computational social science community.

Verification and Validation in Cognitive Modeling and Simulation

In contrast to the computational social science community, verification and validation in cognitive modeling and simulation has longer roots, primarily in the testing and evaluation of modeling and simulation in artificial intelligence (AI), expert systems, and artificial neural networks (Menzies 2002; Menzies and Pecheur 2005; O'Keefe and O'Leary 1993; Skias 2006). Validation of machine intelligence has been a core question for computational models of human reasoning since the field's inception. One of the earliest identified tests to address the validity of an artificial intelligence was the Turing test, proposed by Alan Turing in his paper "Computing Machinery and Intelligence" (Turing 1950). Turing famously proposed an "imitation game" involving two human beings and a machine-based intelligence. One of the human beings, the questioner, would "converse" simultaneously with both the machine and the human, but not know which entity was answering which questions. A machine intelligence that "passed" the Turing test would be one that was so fluent in answering the questioner's queries that the human questioner could not tell the difference between the machine responder and the human responder. In fact, later in this paper, we discuss a modification of the Turing test – a CAPTCHA – as a possible evaluative approach for cognitive models.

More recently, verification and validation researchers working on the evaluation of artificial neural networks and expert systems have adopted a more integrated, software lifecycle approach to evaluation. As Menzies and Pecheur write, "AI software is still software, albeit sometimes quite complex software. Hence, methods developed for assessing normal soft ware systems still apply to [artificial intelligence] systems" (Menzies and Pecheur 2005). For example, standard testing to identify defects, failures and compatibility issues in the code are useful in evaluating artificial intelligence programs, as are techniques of static analysis, model checking, run-time monitoring and theorem proving (more generally so-called formal methods). Artificial intelligence software, such as that used in control systems for space vehicle navigation, can be monstrously complex and this complexity creates challenges for all software evaluation techniques that are software engineering based. Menzies and Pecheur write that AI systems can be complex, declarative and model-based, nondeterministic, even adaptive. Each of these characteristics implies different approaches for assessing the "goodness" of the software system for its intended purpose. Some techniques from software engineering are useful in reducing complexity; however, evaluative approaches need to be considered prior to the design and construction of the system: for example, adaptive software is capable of building or tuning its responses based on inputs from the external environment. Verifying and validating that the software is correctly doing what it was designed to do, and that the design is correct in some sense, probably requires access to several training and testing sets; for example, to apply methods like N-way cross validation to determine if a model learned against one domain is valid against data not used in training (Menzies and Pecheur 2005).

Verification and Validation in Computational Science and Engineering

As discussed above, the concepts of verification and validation in computational cognitive and social science refer broadly to concepts and techniques aimed at assessing the correctness of a simulation. Verification an validation are often used interchangeably; for example, to refer to processes intended to assess the correctness of a model vis-à-vis the real world phenomenon it is intended to represent. In contrast, in computational science and engineering, the approach to model evaluation tends to be far more structured (Roache 1998). As Mackenzie points out, demonstrating the dependability of computer hardware and software has been a problem for the engineering community since the 1960s, when computers began to be incorporated as control systems for complex machinery (Mackenzie 1995; Mackenzie 2001).

One of the seminal papers that shaped the direction of verification and validation research in the computational science and engineering community was Robert G. Sargent's paper, "Validation of Simulation Models" (Sargent 1979), see also (Sargent 1985; Sargent 1987). Sargent emphasized that "model validation is not completely separable from model building… that validation techniques should be used during the model building process and also that model building is an iterative process, and confidence in the model is increased from model iteration to model iteration" (Sargent 1979). Sargent developed a framework for relating verification and validation activities to the modeling process, as captured in the "Sargent Triangle," (also called the Sargent Circle) pictured below:

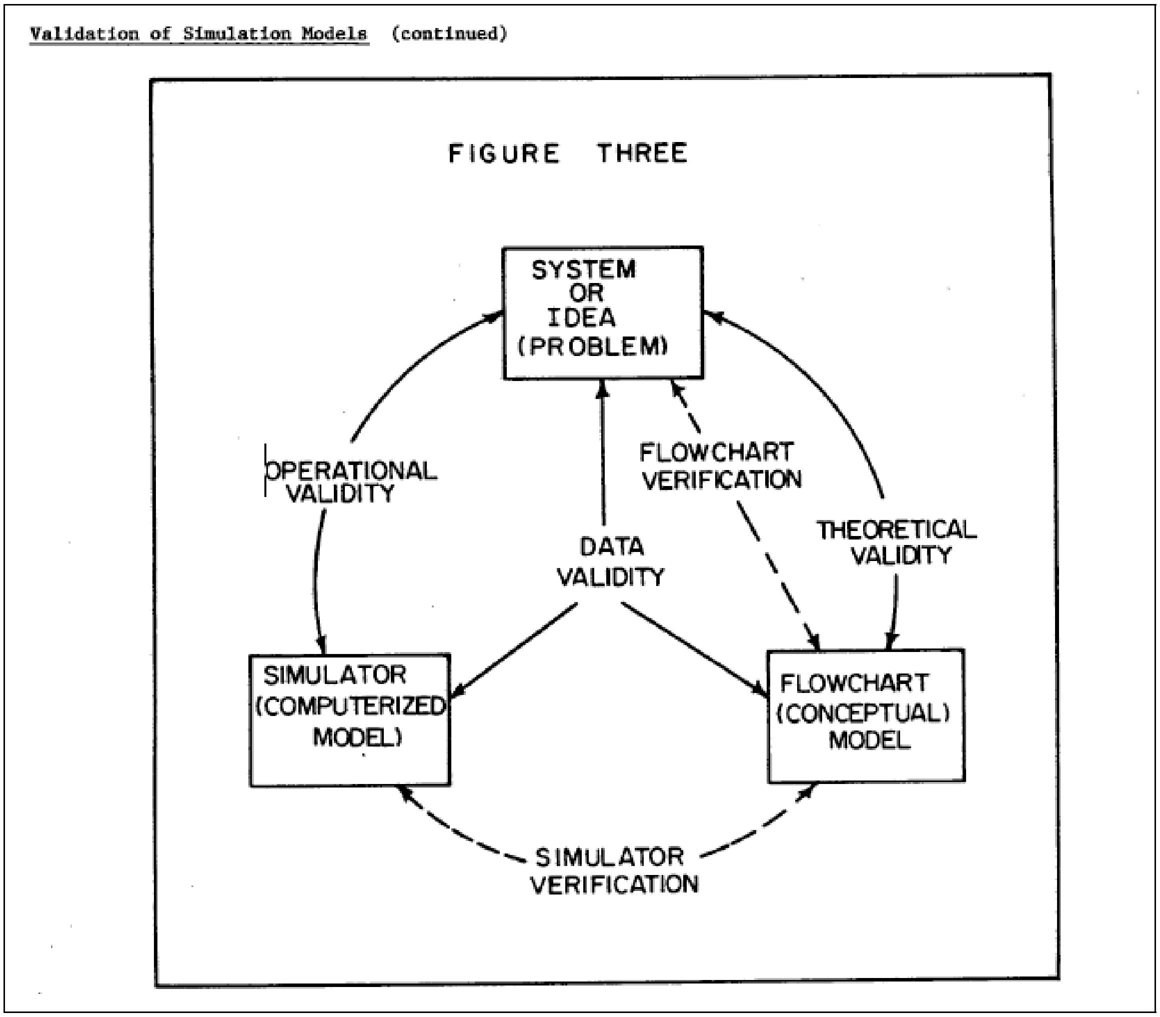


Figure 1: Sargent Triangle (from Sargent 1979)

Sargent's triangle illustrates the relationships among the real-world problem area or phenomenon of interest; the conceptual model of that phenomenon, and its instantiation and operation in code (software). Exercising these relationships in a way that users and stakeholders can understand – as communicated through documentation, stakeholder participation, or both – is what builds credibility in a simulation's outputs. However, Sargent emphasized the importance of data validity as centrally important in building credibility of a simulation among users. A lack of valid data, Sargent wrote, "is the most common reason that initial attempts to validate a model fails (sic)" (498). As we point out in this paper, getting the right data at the right level of quality to support model validation presents an important planning problem for the simulation's stakeholder community.

In computational science and engineering modeling and simulation, verification and validation are typically defined as follows:

- *Verification* – the process of demonstrating on the basis of evidence that the underlying mathematical equations are solved correctly.

- *Validation* – the process of demonstrating on the basis of evidence that the underlying mathematical equations are correct.

It is important to understand that the meaning of "correct" in the definition of verification is operationally interpreted to mean "accurate solution." "Accuracy," of course, is a relative concept; and in this case, it is relative to the intended application of the model. This concept encompasses evidence that demonstrates mathematical and software correctness underlying any evidence of solution accuracy. In other words, the illusion of apparent computational accuracy may be worthless if it rests on a large accumulation of algorithmic and software errors. This may seen contrary to common sense, but consider for a moment how difficult it is to understand whether a complex computer calculation is mathematically accurate – hence the importance of verification. The meaning of "correct" in validation is more problematic. Validation involves the comparison of a verified simulation's results against comparable experimental data, as a means of providing evidence that the equations are more than just *useful* for the intended application. Demonstrating evidence of utility says very little about whether or not the equations actually represent a correct abstraction of the real phenomenon. However, in most computational science and engineering, one could invest a nearly infinite amount of resources demonstrating that a model is true in an absolute sense. Validation, then, is about evaluating whether the model, as specified in the equations and implemented in software, is a correct approximation of the phenomenon of interest for the intended application. The heart of the technical matter lies in understanding what "evidence" means, how to accumulate it, and how to properly use it. We will repeatedly return to the issue of ***referents,*** or clearly specified and meaningful points of comparison that enable us to assess a simulation in relation to reality, that are required as part of the accumulation of evidence. Generally, these referents are mathematically centered for verification, and physically (that is, experimentally or "observationally") centered for validation.

The most canonical example of V&V in computational science and engineering centers on the solution of various partial differential equations of "mathematical physics," which is the main focus of the National Nuclear Security Administration's Advanced Simulation and Computing (ASC) program (Kuznezov, 2004).

Our emphasis on conceptual models in these definitions derives from V&V guidance developed for human interaction models in the Defense Modeling and Simulation Organization (DMSO), which – to our knowledge – is the only institutionalized V&V program that directly addresses the problem of modeling and simulation for human (social and psychological) phenomena (see for example (Davis 1992). Other elements of our guidance derive from work done under the ASC program in the Department of Energy's National Laboratories (Ang, et al. 1998; Christensen 1998; Cornwall 2005; Futral, et al. 1999; Hodges, et al. 2001; Kerbyson, et al. 2005; Logan and Nitta 2002; Nitta and Logan 2004; Nowak and Christensen 1997; Pilch, et al. 2001; Post and Kendall 2004; Trucano 2005; Trucano, et al. 2001; Trucano, et al. 2002).

Professional Standards for Verification and Validation

Because modeling and simulation software is so widely used in the engineering community, software evaluation has been an increasingly important topic for the professional societies that set standards for industry practice. Since the late 1960s, several professional engineering and computer science associations, as well as major federal R&D programs, have established and promulgated formal and consistent definitions for verification and validation, in addition to standards for planning, conducting, and documenting evaluation activities as part of a large lifecycle approach to software development and usage. For example, the Institute of Electronics and Electrical Engineers (IEEE), which works to develop and promulgate engineering standards internationally, treats verification and validation as "a disciplined approach to assessing software products throughout the product life cycle… [and] employs review, analysis and testing techniques to determine whether a software system and its intermediate products comply with requirements." This systematic approach to evaluation as an embedded activity in a software's lifecycle is reflected in IEEE Standard 1012-1986, *IEEE Standard for Software Verification and Validation Plans*, which uses the following definitions for verification and validation:

- *Verification:* The process of determining whether or not the products of a given phase of the software development cycle fulfill the requirements established during the previous phase.
- *Validation:* The process of evaluating software at the end of the software development process to ensure compliance with software requirements.

Other professional societies – including the American Society of Mechanical Engineers (ASME) and the American Institute of Aeronautics and Astronautics (AIAA) – have also established standards for Verification and Validation. Both ASME and AIAA treat verification and validation as an iterative process embedded in the software development lifecycle, from requirements through design, coding, and software testing. AIAA also sets out guidance for assessing the modeling and simulation software in its *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*, AIAA G-077-1998 (American Institute of Aeronautics and Astronautics 1998), which presents definitions for V&V that are similar to those that IEEE provides:

- *Verification:* The process of determining that a model implementation accurately represents the developer's conceptual description of the model and the solution to the model.
- *Validation:* The process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model.

Lastly, we emphasize that there is no such thing as a verified and validated simulation "code" or software package. Instead, verification and validation activities should focus on the intended application of the code for a specified problem area. This is because most modeling and simulation tools can be applied in a range of problem spaces, which makes for substantial variability in the selection and construction of input parameters and outputs. Moreover, codes are living artifacts whose functionality evolves as people engage with the code's structure: adding new algorithms, updating existing ones, and addressing bugs. Given the indefinite possibilities for applying a code, and the likely evolution of its structure with successive releases, a stamp of "Verified and Validated!" makes little sense and is probably misleading. Instead, verification and validation are ongoing activities that provide at least a minimal evaluation function for specified applications of the code, above and beyond “normal” software quality management practices (e.g., requirements documentation, version control, regression testing, bug tracking).

Institutionalizing Verification and Validation

Both AIAA and IEEE approach verification and validation as part of a larger, more extensive software lifecycle management problem (in some ways only implicitly), a context in which verification and validation moves from the realm of ad hoc methods and applications to formal programmatic investment. In this context, V&V can quickly become an expensive and time-consuming process. As such, it is perhaps not surprising that the most formal investments in verification and validation programs are made primarily by large government institutions with big budgets: for example, the National Aeronautical and Space Administration (NASA); the Department of Defense (DoD), the Department of Energy (DOE) and the Nuclear Regulatory Commission (NRC). Each of these organizations has historically relied heavily on software for modeling and simulation and/or technology operations. The engineered systems they build and run are prone to high-visibility, high-consequence (even catastrophic) failure that can entail loss of expensive equipment, facilities, and programs, not to mention human lives, and public trust. Hence, there are strong rationales and requirements for formal evaluation of modeling and simulation codes whose outputs are intended as information for high consequence decisions.

An Example of an Institutionalized Program: Verification and Validation at NASA

At NASA long history of relying on software to perform critical guidance and navigation functions drove the agency to invest significant resources in formal processes for evaluating software performance and reliability. One of NASA’s earliest brushes with software failure was in 1969. An error in the gravitational model being used by the Apollo 11 Lunar module’s control system nearly sent the module into a boulder. Astronaut Neil Armstrong saved the day when he assumed manual control of the spacecraft and guided it to a safe place (Mackenzie 2001). By the 1980s, NASA had poured millions into developing the painstaking process of debugging and “proofing” code as part and parcel of the development process. However, the 1986 Challenger accident (which was attributed to a faulty O-ring, not a software glitch) drove NASA to invest more resources throughout its programs in mission assurance, risk management and safety, including computater hardware and software. This led to the formation of an Independent Verification and Validation program (IV&V), including the construction of a dedicated facility in West Virginia. Over 150 staff members work with partners in industry and academia to develop software assurance methods for mission critical applications. To deliver more objective analysis, verification and validation activities at NASA are conducted independently of the software development team. This means that the V&V program prioritizes its efforts, submits independent reports, and maintains a separate budget from the software development team (for more information, go to http://www.nasa.gov/centers/ivv/home/).

THE DEPARTMENT OF ENERGY, STOCKPILE STEWARDSHIP AND PREDICTIVE COMPUTING

Since 1992, the United States has been observing a moratorium on underground nuclear tests. At the same time, the US has also significantly cut back the size and diversity of the US nuclear stockpile, and reduced its investments in above-ground testing activities. This creates a context in which alternative experimental, inspection, and computational methodologies are critical for evaluating the security, safety, and reliability of an enduring nuclear stockpile; and in which computational modeling and simulation represents a significant integrative and predictive technology. One major investment was the Accelerated Strategic Computing Initiative, or ASCI (which we will refer to by its current name, the Advanced Simulation and Computing program, ASC), which represented a multi-hundred-million-dollar investment in scientific computing. ASC focused on developing a new suite of computational physics and engineering codes and high performance, at-the-frontier computing hardware for running high-fidelity simulations of the physical processes that occur in nuclear weapons. Verification and validation were not formal project elements of the ASC program when it was implemented.

However, by 1997, DOE NNSA leadership recognized the need for a formal evaluation program, given that the codes were to play a main role in supporting decisions related to the certification of the nuclear stockpile. The vision of the ASC V&V program was to "establish confidence in the simulations supporting the Stockpile Stewardship Program through systematic demonstration and documentation of the predictive capability of the codes and their underlying models" (Trucano and Moya 2003). Since 1999, the ASC program has invested significant resources in verification and validation, and has developed a widely respected verification and validation program, one that has both drawn on and enhanced standards set by other professional bodies (e.g., AIAA, NASA, DoD).

Verification and Validation in Sandia's ASC Program

Within the DOE's Advanced Simulation and Computing program, Sandia National Laboratories has played a key role in developing the verification and validation elements for Department of Energy's ASC program. The program is operationalized around the definitions from the Department of Energy's National Nuclear Security Administration (NNSA):

> *Verification* is the process of determining that a computational software implementation correctly represents a model of a physical process
>
> *Validation* is the process of determining the degree to which a computer model is an accurate representation of the real world from the perspective of the intended model application.

Several key themes have remained constant in Sandia's V&V efforts since the inception of the ASC V&V program, and these themes are entirely relevant and appropriate for computational social and behavioral models as well. Firstly, verification and validation go hand-in-hand. Verification is critical because it provides a basis for validation, insofar as it is impossible to judge the goodness of a code's output without knowing if the simulation software properly implements the conceptual model. Conceptually, verification provides information that is a necessary condition for the correct design, execution, and interpretation of validation (experimental-based) activities. At the same time, verification tells one little about the code's performance in the absence of validation; on its own, verification is logically and philosophically incomplete. Therefore, it is fundamental that verification and validation are inextricably coupled in a V&V process.

Secondly, neither verification nor validation is a one-time only activity that a code team has to perform, say at the end of a development project, to demonstrate that they got it (the software) "right." Instead, V&V is treated as a process for evaluating and improving the quality of computational modeling and simulation software. V&V is a hierarchical and continuously executed methodology through which stakeholders acquire evidence about a model's predictive capabilities. These are ongoing activities that begin in the earliest stages of the code project, that are woven throughout the development process, and that are deliberately engaged whenever the software is being directed to a new application area.

Lastly, we believe that the "science" of performing experimental-computational comparisons in "computational science" remains immature (see Oberkampf and Trucano, 2002; Oberkampf, Trucano, and Hirsch, 2003 for recent reviews). An important goal of the ASC V&V program is to advocate work that improves the "science" of these comparisons, and thus strengthens the conclusions that can be drawn from them, especially for projects at SNL that involve high consequence decisions (as in the nuclear weapons programs) which use modeling and simulation output as a inputs into those decisions.

It is essential to emphasize that there is nothing in these three key themes that is actually specific to the subject matter of computational physics and engineering, or to nuclear weapons stockpile problems for that matter. This is one of the important points of our work and our views of what the central core of computational model evaluation activities for computational social and behavioral models should be.

If this seems like a somewhat heavy-handed approach to code evaluation, it is because V&V in the ASC program involves the assessment of a modeling and simulation code's predictive capabilities in regards to decision making under conditions of risk and uncertainty. In other words, *the extremes (as some might see it) that drive the ASC V&V methodology really have nothing to do with subject matter content, and everything to do with the intended (high-consequence) application(s) of the computational modeling*. The importance of the application defines the level of effort expended in model evaluation, not the "quantitative" characteristics of the subject matter area. It is hard to overemphasize the importance of this insight.

In the next section, we explain how model evaluation – V&V – is situated in the larger context of stockpile computing (the intended application), which encompasses research, development, deployment, and application of computational technologies, including hardware, software and associated methodologies, to support high consequence decision-making related to the US nuclear stockpile. Key elements of stockpile computing that generalize to a much broader domain of consequential computational simulation, including computational social and behavioral models, and especially the role of V&V in supporting stockpile computing, are discussed below. We will continue to refer to "stockpile computing" as a brief euphemism for the general challenge of conducting computational simulation in consequential decision environments. The scope of these remarks is quite general, nonetheless.

Verification and Validation in Context

The historical evolution of the ASC V&V program has created a current state in which V&V activities interplay with a set of evaluative methodologies aimed at specifying, describing and using credibility of computational simulation within defined application domains. In this section, we describe the context for ASC V&V efforts, emphasizing the role of modeling and simulation codes in risk informed decision-making. In the context of NNSA "Stockpile Computing," the computational science that supports the management of the U.S. nuclear stockpile, the ASC program creates computational methodologies and technologies, and provides a functional basis for skilled analysts to support high-consequence national security decision-making in the specific domanin of nuclear weapons performance. The success of stockpile computing depends on how effectively computing resources and areas of expertise are aligned with the modeling and simulation requirements for maintaining the US nuclear stockpile, among other factors. Because ASC computational capabilities play such a critical role, they are subject to rigorous evaluation and scrutiny to assess and demonstrate their credibility.[1] It is important to recognize that the ASC programmatic emphasis is on the applications of computational models, so that verification and validation activities provide additional information to people who are responsible for using model outputs in real-world decisions. For example, Figure 2 illustrates how validation activities – in the form of a systematic comparison between experimental and calculation, and in turn a comparison with the requirements for solution accuracy, help decision makers assess whether or not the model is ready for prime time – that is, use in a real-world decision environment. This is a very straightforward logic and it is completely independent of specific physics,and engineering content, or nuclear weapons-specific applications. It exists because of the *intended use of the computational science*.

[1] As Pilch, Trucano, Peercy, Hodges, and Froelich point out, an integrated approach to stockpile computing focuses not just on codes, but also on those "capabilities" embodied in experienced analysts and computational infrastructure. They suggest levels of formality for assessing the qualifications of each of these, depending on the level of risk and consequence involved in the decision area. See Pilch et al 2004: 10-15.

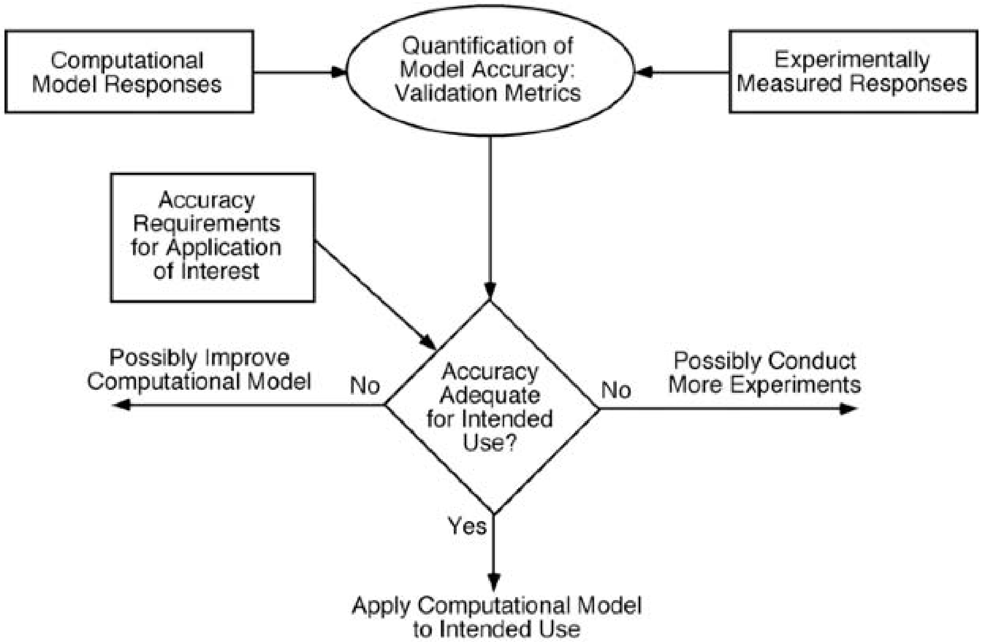


Figure 2: Decision Flow Diagram for Asssessing Adequacy of Model/Simulation for Intended Application (Oberkampf and Trucano 2004)

Uncertainty quantification, analysis, and communication must also be key elements in verification and validation efforts, most basically so that people can take uncertainties into account when contemplating decisions that make use of model predictions. The fundamental strategy to drive uncertainty quantification in model evaluation and application is to have the goal of expressing all model results in the form of a ***best estimate plus uncertainty (BE+U)***. As we discuss below, verification and validation activities do not eliminate uncertainty. Instead, they account for it as thoroughly as possible. QMU is one expression of methodological principles that support the delivery of computational modeling results in the form of BE+U and with appropriate compensation for the decision-making environment that needs the modeling results. The specific context in which QMU has arisen has been stockpile management, but the core principles are independent of this context. As we will mention, QMU itself, as Sandia has argued, is to a certain extent a product of a broader subject matter area in which uncertainty and hich-complexity, high-consequence decisions are oftern present – quantitative risk analysis.

Stockpile Computing, QMU, and the PCMM

Within the Sandia ASC program, verification and validation are part of a larger set of evaluative and decision support methodologies for computational modeling. A flow chart describing the context in which V&V takes place is included below, illustrating the eight different classes of activity that take place in order to evaluate code credibility:

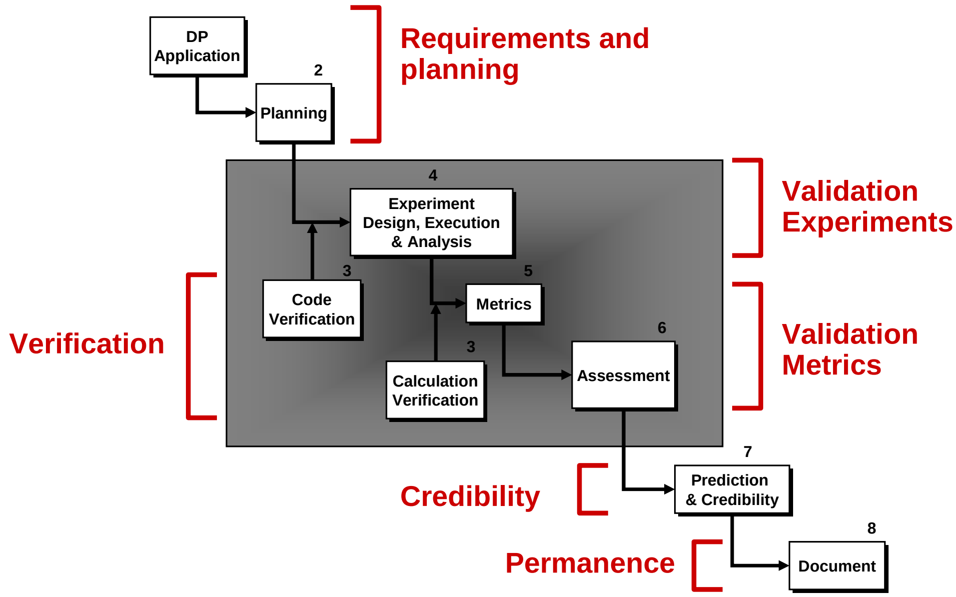


Figure 3: Verification and Validation in Context

The grey box in Figure 3 emphasizes the domain of experimental validation activities and their intersection with verification activities. This process is described in detail in Trucano et al (2002). All of these elements are universal in all applications of computational modeling, especially if one mollifies the apparent emphasis on "validation experiments" to "observational referents." More generally, one should view "observational referents" as tests of the external accuracy of the modeling in comparison with the "real world." When "real world referents" do not exist, and this must be understood in the context of the fully hierarchical validation methodology laid out by Trucano and his colleagues, then a serious question arises about the ability to perform validation at all. In our view this raises serious questions about the limits of applicability of such models, and increases the risk associated with their application. Difficulty in validation of given models thus makes more prominent the necessity and role of risk analysis in applying the models.

Rather than summarize details underlying the proposed validation methodology presented in Figure 3's schematic, we see it to be of particular importance to further explain the overarching context with the three key conceptual frameworks that we have already mentioned: the constituent concepts (or elements) for stockpile computing; the intellectual framework of prediction in the context of quantification of Margins of uncertainty (QMU); and the assessment of predictive capability, specifically the Predictive Capability Maturity Model (PCMM) that is advocated by the Sandia ASC program. We reiterate that our point of view in calling attention to these frameworks is that they serve as a general starting point for our view of what rigorous computational evaluation methodologies that is independent of subject matter area. Each of these is discussed briefly below, and the reader should keep our relatively domain-independent perspective in mind.

Stockpile Computing Concepts

We very briefly comment on this framework, which is discussed in great detail in Pilch et al (Pilch, et al. 2004). Simply stated, the stockpile computing concepts framework addresses all the components that contribute to the formulation, execution, analysis, communication and application of computational modeling. From the perspective of model evaluation for consequential applications, the fundamental principle instantiated in this framwork is that the computational modeling is a chain of key activities, and that the final results are only as strong, for example from the perspective of credibility, as the weakest link in that chain. One result of this kind of thinking is that the model itself is only one of the elements that produces computational results that may be used or considered by decision makers.

The framework basically articulates a modeling process consisting of several elements, and an assessment framework that can be used to measure the quality of each of the elements. From a high level, the elements that are analyzed and considered for quality assessment in Pilch et al (2004) are: (1) the input information that is required for performing modeling; (2) the model itself; (3) the model results and their transformation through post-processing; (4) the infrastructure that is necessary for creating and using input and output, and that is necessary for executing the model; (5) the human being(s) who conduct the modeling.

For example, by input information we mean geometric specifications and physical parameters required for computational physics simulations; more general notions of "input" are included in this category for social and behavioral models. What we mean by the model itself is self-evident and is transparent independent of subject-matter domain. The third element emphasizes that most modeling results are rarely "the numbers" that emerge from a calculation using the model. Rather, model "results" are actually transformed using a variety of tools and techniques into other "results." This transformation is accomplished most familiarly with visualization tools. But an equally important, and more complex transformation, could be statistical processing of results, as would be highly likely in social and behavioral modeling. Infrastructure is also a relatively transparent concept across subject domains. It refers to the computing infrastructure that enables computation, which includes communications, storage systems, the computing hardware that models are executed upon, and the software that underlies these systems. The fifth element, that is the capabilities and skill levels of the humans that are performing the work, may be seen as controversial, but we consider it to be a necessary element for evaluating the quality of the chain leading to modeling results.

As well as a detailed analysis of the role of these elements in producing modeling results that may be appropriate for consequential applications, Trucano et al (2004) also consider evaluation strategies for assessing the quality of these elements. Specific to the concerns of this paper, it is obvious that "model evaluation" (or V&V) is the essential means of evaluating element (3). From the more general perspective of this framework, V&V is also relevant to all the software that creeps into the modeling chain, including pre- and post-processing and infrastructure. Broader issues of evaluating the quality of input, post-processing, infrastructure, and human factors immediately arise, however, and much of our consideration of stockpile computing concepts centers on how to make sense of this diverse set of elements and aggregation of their quality assessment.

The description above basically lays out key elements of the modeling chain that have to be of concern because of their influence on credibility for any subject matter domain. The degree to which these elements are rigorously scrutinized is almost entirely dependent upon the rigor of the intended application. This is a framework that is germane to social and human interaction modeling as well as to computational physical science in nuclear weapons applications. As is the case for all of the frameworks we discuss, an understanding of the cost-benefits associated with implementation of the framework is a realistic concern. Implementation of the stockpile computing concepts framework centers on whether the framework provides value for whatever level of cost is expended. This is not a subject matter area issue, but rather an application-specific issue. We do not address this issue in the broad content of this paper for the simple reason that it can best be addressed in the context of specific modeling projects.

Quantification of Margins and Uncertainties (QMU)

QMU is one example of a framework for performing uncertainty quantification, linking it to rigorous assessments of modeling credibility, and communicating and applying this information in consequential decision environment where risk must be managed. The relevant reference on this topic for purposes of this report is that whitepaper of Pilch, Trucano, and Helton (2006). Uncertainty quantification is an important issue and decisive methodology in model evaluation, especially validation (Trucano et al. 2002). The details of how to do it, why to do it, and how to use it are sensitive to subject-matter domain, but the methodological need for involving uncertainty quantification in model evaluation is domain independent. In fact, we point out that the role of uncertainty quantification in model evaluation is more complex and more essential in social and behavioral modeling than in physical science, one reason being the larger magnitude and diversity of uncertainty in observational data.

QMU pursues uncertainty quantification in a quantum jump beyond the needs of model evaluation. Fundamentally, QMU is one way of expressing a broader concern the role of modeling in risk-informed decision making. Such decision-making must not only deal with the risk embedded in the fundamental decision context, for example the safety of nuclear reactors and the creation of technically rigorous and effective regulation of these systems, but must also deal with risk associated with the credibility of the modeling. In such an environment, simply performing model evaluation does not necessarily mean that these assessments will be useful within the decision-making environment. A rather different set of considerations must be performed to guarantee this. The whitepaper of Pilch et al. analyzes this consideration and provides an extensive set of references linking the issues of model evaluation and risk assessment and management. This analysis is effectively independent of both the physical science domain and the nuclear weapons context emphasized in the Pilch etal. report.

Specific to the nuclear weapons context in which QMU has emerged, particular notions about nuclear weapons performance, thresholds of acceptable performance, distance of believed performance from these thresholds – called ***margins*** – and ***uncertainty*** in all of these concepts have been defined and analyzed. All of these concepts will vary depending on the underlying subject matter of the modeling and on the intended application. But the concept of risk, the components of risk that are *introduced* by the modeling, the *uncertainty* in model results as well as *uncertainty in the credibility* of those results, and the need to *inform decision making* performed in the presence of this risk transcends specific domain considerations, and centers on the nature of the application of the modeling. We believe that the general principles underlying Pilch et al. are important for social and behavioral modeling applications in consequential environments, and are certainly necessary for placing model evaluation activities in the proper context of consequential modeling applications.

The Predictive Capability Maturity Model (PCMM).

The PCMM framework (Pilch et al. 2006, Oberkampf et al. 2007) creates a methodology and schema for quantitatively measuring and assessing the progress of a modeling and simulation software application toward predictive capability. It also supports assessment of an application's readiness to support levels of decision making under increasing conditions of risk and consequence. The PCMM methodology uses a two-axis matrix that identifies the key components of predictive capability, and then assesses requirements for these components under increasing conditions of risk.

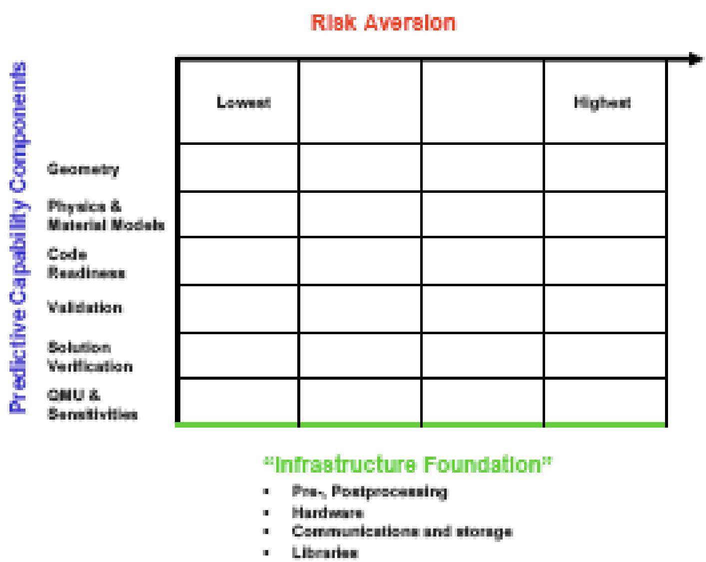


Figure 4: The Predictive Capability Maturity Model (Oberkampf, Pilch and Trucano, 2007)

The PCMM supports generation of a relatively complex multi-dimensional metric for evaluating necessity, progress, and sufficiency of M&S capability for intended applications. The PCMM is currently being used operationally in the Sandia ASC program to guide V&V, model capability assessments, and computational model peer review activities. This implementation is expected to grow over the short-term. The experience generated from this implementation will likely have much of interest for sharpening our understanding of the assessment of predictive capability of social and behavioral models as well.

In our view, the emphasis on "prediction" (and we avoid a jargon-ridden discussion of exactly what this word could mean in every conceivable context) is entirely appropriate for social and behavioral modeling. The main reason we claim this is that, once again, almost all consequential applications of these models are engaged in some form of prediction. It seems self-evident that being able to create rigor and standardization around the assessment, communication and comprehension of predictive capability of these models is highly desirable. This framework seems to be especially important for consequential applications of social and behavioral models.

Verification and Validation Activities in a Stockpile Computing Context

In a stockpile computing context, the process diagram in Figure 3 ordering the execution of V&V activities is not just a basket of evaluative methodologies. Rather, a programmatic commitment to formal V&V has a ripple effect throughout the entire ASC program, as it generates requirements for code project planning, staffing, and funding; as well as code development, testing, documentation, and experimentation. Without adequate planning, it is difficult to ensure the requisite accumulation of quality evidence that will demonstrate the code's credibility in relation to a particular application area. As such, and as characterized in Figure 3, verification and validation involve a great deal of planning, including the identification of stockpile drivers, planning for V&V activities, code development, software, algorithm and solution verification activities; validation experiments; and uncertainty quantification. We discuss each of these briefly below.

- **Stockpile Drivers**. (In general, for general modeling these are simply application requirements that underlie given modeling efforts. They are called "Drivers" in the ASC V&V context because of the philosophy of that program that all the work is fundamentally "driven" by these requirements.) V&V begins, or should begin, with a request from the stockpile community for computational analysis. At this point, the problem owner (usually a representative from the Directed Stockpile Work Campaign in the nuclear weapon context) works with the ASC program to identify specific stockpile drivers, or application areas where modeling and simulation are required. Identifying drivers requires extensive elicitation and documentation of needs, requirements, and scenarios that help define the needs for computational analysis, requirements for the computational analysis, including how the results should be delivered, accuracy, and documentation; and even the calculations to be performed.

- **The Phenomenon Identification and Ranking Table** (PIRT). To translate stockpile (or different) needs, requirements and scenarios into specific computing requirements, the ASC V&V program relies on a structured planning tool known as the Phenomena Identification and Ranking Table (PIRT). The PIRT is a planning methodology adapted from the nuclear reactor safety community. This is a planning tool used to prioritize verification and validation activities, and to link those with experimental programs in the nuclear weapons program. By extension, the PIRT supports resource allocation decision-making insofar as it identifies the verification and validation activities that are both necessary and sufficient to evaluate the simulation code's credibility. The PIRT structure is summarized in Figure 5. The PIRT characterizes the constituent phenomena of the problem space; ranks them in importance in relation to what the modeling and simulation code is aimed at addressing; and assesses the current state of the model's validity for the identified phenomena (see especially (Pilch, et al. 2004; Pilch, et al. 2001). This information, in the form of a gap analysis, can be used to prioritize validation (hence verification) activities. More generally, the PIRT is a useful way of expressing key modeling elements, is an expression of underlying conceptual

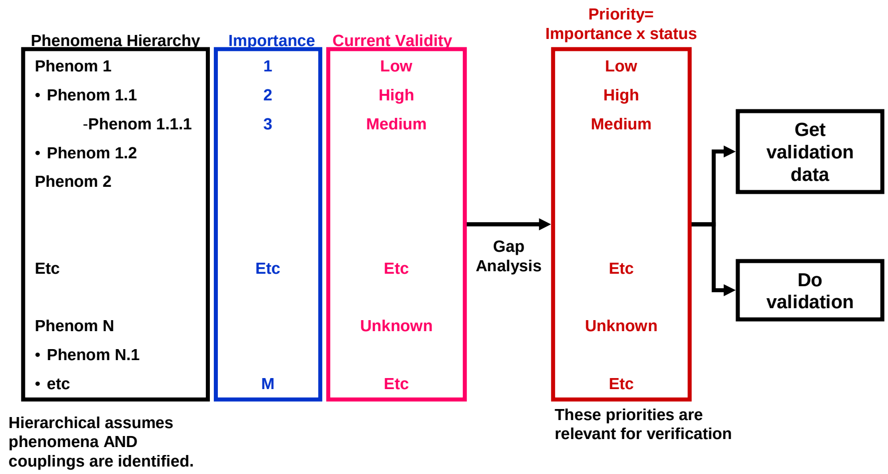


Figure 5: Phenomenon Identification and Ranking Table

- **Software, algorithm, and solution verification.** Verification is a multi-activity process that evaluates the correctness of a code mathematically, algorithmically, and as a software product. Code verification "provides the necessary foundation for believing accuracy assessments" (Klein, et al. 2006) and provides evidence of code correctness, minimizing the possibility that good comparison between a validation calculation and experiment is actually due to the presence of a bug in the code. Verification is at its heart a software-centric activity and therefore transcends particular subject-matter disciplines. All software needs verification evidence.

    - ***Code Verification*** includes activities directed toward finding and removing mistakes in the source code; finding and removing errors in numerical algorithms; and improving software using software quality assurance practices. Good software quality engineering practices support the development and demonstration of code with minimal errors; as do the construction of well-structured test suites, and the selection of assessment criteria to decide if a code has passed a test problem.

    - ***Solution Verification*** emphasizes activities directed toward assuring the accuracy of input and output data for the problem of interest; estimating the numerical solution error (e.g., error due to finite difference/finite volume/finite element mesh resolution, temporal discretization, finite iterations, etc); and assessing how mathematically accurate a given calculation is. Extensive discussions of verification activities can be found in multiple ASC publications (Ang, et al. 1998; Oberkampf and Trucano 2002; Oberkampf, et al. 2004; Oberkampf and Trucano 2006; Trucano and Moya 2003; Trucano, et al. 2002). Assessment of calculation accuracy, even the meaning of "accuracy," is a highly domain dependent concept. We consider a proper extension of conventional concepts of accuracy arising in computational physics and engineering from the solution of partial differential equations to (possibly) very different

- **Validation calculations and experiments.** Validation and prediction are inherently intertwined for consequential modeling applications, and validation work, for example as formulated within ASC, aims at demonstrating some level of predictive capability. This means that validation activities require comparison between a code-generated prediction and the comparison of that predictive calculation to a real-world experiment. As such, planning plays a particularly important role in validation, since both calculations and experiments require precise specification prior to execution to ensure alignment between the two activities. Indeed, validation logically follows from a validation plan that unfolds a strategy that is dominated by the transition from a PIRT (Phenomena Importance and Ranking Table), to a prioritization of verification and validation tasks, to a specification of needed validation that will be performed, and finally to actual validation calculations. This also requires careful analysis and documentation of both experimental and calculation uncertainties involved in the comparison.

    - ***Validation calculations*** are calculations that are compared with experimental data for the purpose of inferring physical quality (physical accuracy) of the associated calculations. Validation calculations have the specific purpose of enabling an assessment of the physical quality/physical accuracy/predictive capability of the code for the application represented by the chosen validation data.

    - The ***experimental data*** that validation calculations are compared with must have specific characteristics in order to be effective in enabling validation (Trucano, Pilch and Oberkampf 2002). Experiments and tests can be conducted for multiple purposes, as argued in Figure 5. Not all experimental data can be considered to be useful for validation. Furthermore, not all comparisons with appropriate experimental data can be considered to be validation in the precise sense that is defined by the ASV V&V program. Experiments and the anticipated application of the experimental data must be carefully designed (using appropriate guidelines from statistics and experimental design) and relevance, especially if they are to be used optimally for judging the credibility of a computational model. The subject code should be engaged in the definition and design of the experiments, not only in the analysis of their outcomes. Oberkampf and Trucano (2004) emphasize the need for objective validation metrics to compare calculation to experiment, as opposed to what they call the "viewgraph norm," in which a visual comparison of graphically-depicted results provides the basis for assessing alignment. Trucano, Pilch and Oberkampf (Trucano, et al. 2002) provide a detailed description of the planning, execution, and comparison of validation calculations and corresponding experiments.

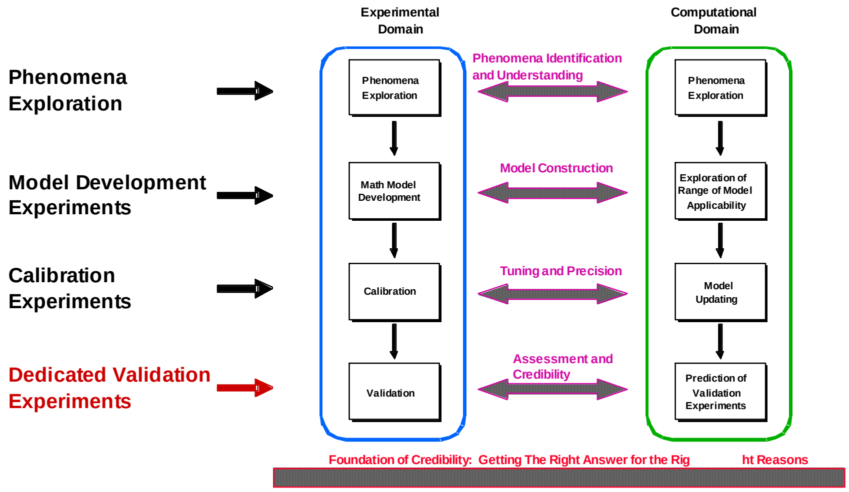


Figure 6: Multiple purposes for experiments. Validation experiments are the most stringent.

- **Uncertainty Quantification.** The ASC program's commitment to the Quantification of Margins of Uncertainty (QMU) framework means that judicious and thoughtful quantification and analysis of uncertainty plays a critical role in assessing whether or not a model/simulation is appropriate for application in a high-consequence decision environment. As noted above, generating a credible comparison between experiment and calculation depends on whether or not uncertainty is systematically accounted for in both activities. Comparisons of calculations and experiments for the purpose of validation require a precise understanding of the presented comparison, which is typically in the form of plots, but could also be detailed tabular comparisons or other quantitative representations of the comparison. In particular, this means that the *uncertainty in the experimental data* and the *numerical accuracy of the presented calculation(s)* must be acknowledged and accounted for in the details of the comparison. Addressing uncertainty in calculations can be computationally challenging, as it requires identification of multiple sources of uncertainty (e.g., in input quantities, model form), the use of appropriate quantification techniques for different classes of uncertainty (*epistemic* or lack-of-knowledge uncertainty. and *aleatory* uncertainty, which points to inherent stochasticity); and the propagation of uncertainty through a calculation. Sensitivity analysis identifies how the uncertainties in different inputs drive uncertainty in the model outputs. A similar systematic approach to documenting and probably minimizing experimental uncertainty supports a sound comparison between experimental and calculation results.

Aligning V&V with Computational Science Software Development

As we discussed earlier, Sargent's triangle provides a conceptual model for identifying classes of activities (verification or validation) for evaluating different aspects of a modeling and simulation code. However, as a description of how computational modeling and simulation codes are actually developed, the Sargent triangle does not come close to capturing the complexity of development activities involved in R&D codes. Post and his colleagues (Post and Kendall 2004; Post and Votta 2005) have analyzed this issue, and observe that the development of modeling and simulation codes (computational science software) is a complicated, interdisciplinary, iterative process that typically involves multiple parallel and sequential activities being pursued by different team members with different disciplinary backgrounds. Standard software lifecycle models often fail to correctly address code development issues in computational science, both because of both the unusual content in computational science software and because of the often open-ended environment in which computational science codes are developed. Figure 7 is a representation developed by Post et al. to illustrate the "real" workflow that characterizes a computational science project. As this diagram indicates, computational science R&D codes – most emphatically including those in social, cognitive, as well as physical, or engineering sciences – involve complex work flow cycles, rather than relatively serial life cycle characteristics. From our perspective, this diagram does a far better job capturing what really happens as a modeling and simulation team embarks and progresses on a code development effort than do most conventional software lifecycle engineering rubrics.

Our concern over the proper lifecycle for, say, computational social and behavioral models that we may be interested in evaluating, is that the lifecycle structure intersects the design and execution of V&V, and at least partly influences cost-benefit considerations associated with V&V. Verification and validation methods and implementations that do not take the complexity of computational science R&D projects into account are destined to fail. Aligning the principles of the Sargent triangle with the realities of a modeling and simulation software project is no small task. For that reason, we will repeatedly emphasize the importance of planning in developing realistic and successful verification and validation methodologies.

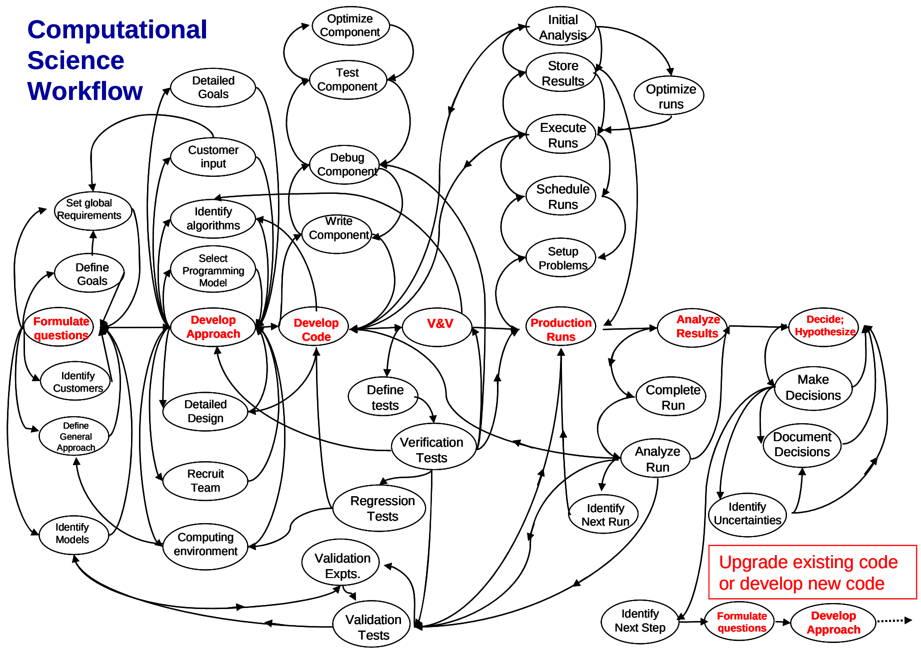


Figure 7: Computational Science is Complicated

VERIFICATION AND VALIDATION IN COMPUTATIONAL COGNITIVE SCIENCE: A FRAMEWORK FOR MODEL EVALUATION

In this section of the report, we present a framework for addressing issues of computational model/simulation evaluation in cognitive science and technology. In doing so, we build on the ASC verification and validation approach described above. However, we have broadened the approach to include several additional activities that address some of the issues that differentiate simulation construction in the physical sciences from the social and cognitive realms. We have tried to organize these activities so that they support the ongoing development of the modeling project, adding as much value as possible to both the process and the outcome.

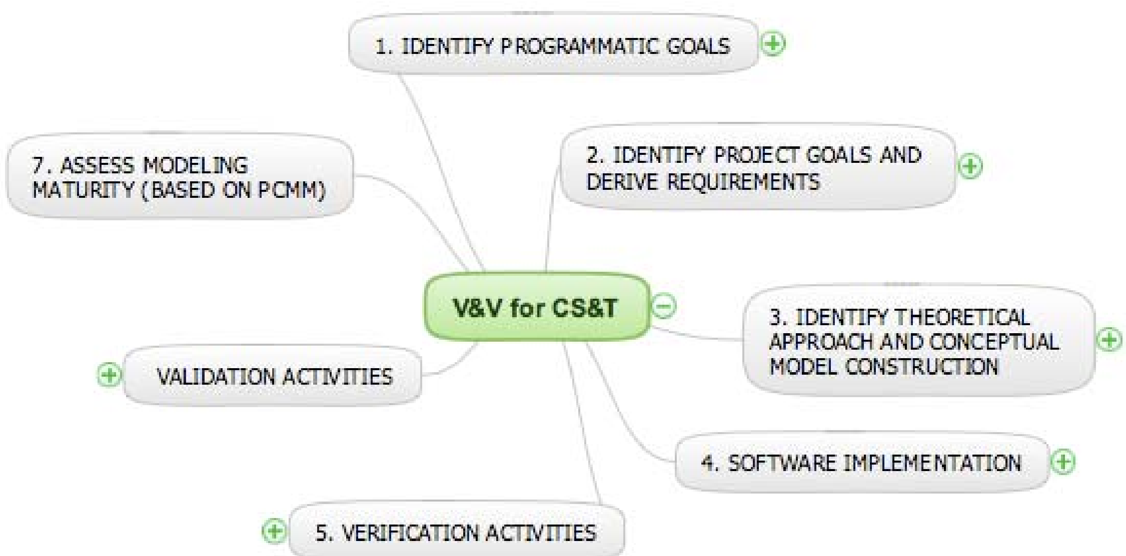


Figure 8: V&V for CS&T - a Basic Framework

The approach we describe in the following pages consists of seven areas of activity that are aimed at building transparency into model and software development, and at systematically developing a base of information that can be used to evaluate the performance of the modeling and simulation project against the goals that its proponents set out. These activity areas include the identification of programmatic goals; identification of project goals and derivation of requirements; identify theoretical approach and conceptual model construction; implementation in software; verification activities; validation activities; and assessment of modeling maturity. As we discuss below, we have adopted two key ASC frameworks – the PIRT and the PCMM – for application in this realm. The other relevant frameworks that we discussed above, a "stockpile computing" framework and QMU, are also appropriate (as we claimed) but beyond the scope of this phase of our work.

A Word on Extending CS&E V&V to Computational Cognitive and/or Social Science

In our study, we created analogs of ASC V&V concepts for application to cognitive science and technology computational models. The logic of the components for assessment that have evolved in the ASC program rests fairly rigidly on the concept that ASC M&S is ultimately about prediction and decision support, in particular for the U.S. nuclear weapons program. We carry this logic over strictly for cognitive and social modeling, and emphasize that V&V assesses models in terms of their value for decision support. A conceptual mapping of ASC V&V tools and approaches from CS&E to CS&T is pictured in Figure 9. The reader is invited to contrast this logic diagram with that presented in Figure 3. The similarities are large, but we have placed some emphasis on the three frameworks that we discussed earlier, plus customized for CS&T the positioning of some of the elements. (The specific numbers indicated in legends in this figure point at specific phases of CS&T software development that are presented in the M&R project design document. These are thus specific links to M&R.)

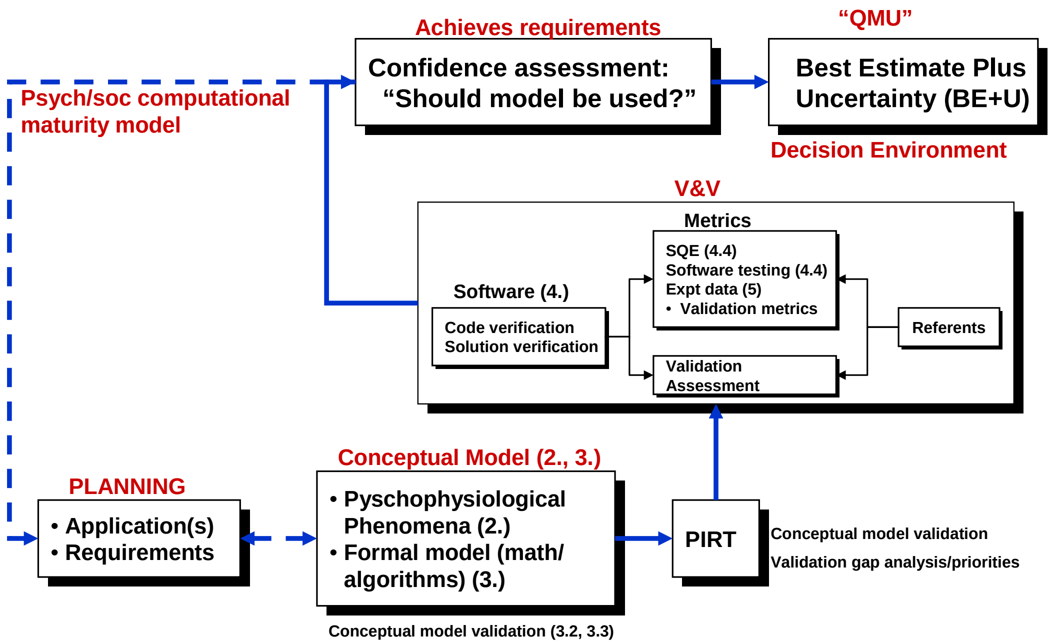


Figure 9: Application of ASC V&V tools and approaches to Cognitive/Social Modeling and Simulation V&V

We recognize that cognitive modeling and simulation software is quite different from codes that simulate the performance of mechanical or electrical engineered systems. Over the course of this project, we have heard on numerous occasions comments along the following lines: "It's impossible to validate models of social phenomena because you can't conduct experiments/can't demonstrate predictive accuracy/can't gather validation quality data!" (Similar arguments have been made in other fields besides social and cognitive science; see for example (Konikow and Bredehoeft 1992). We believe these comments misunderstand the larger purpose of verification and validation: not to prove a model "correct" in an absolute sense, but to support the accumulation of evidence that a model is appropriate for the use for which it is intended.

Much of the epistemology, methodology, and even specific approaches to verification and validation that have been developed for computational science and engineering can be applied in simulations of social, cognitive, or biological systems. For one thing, all modeling and simulation efforts, regardless of discipline, share similar high-level characteristics that predispose them to systematic evaluation, both as software and as attempts to represent real-world phenomena using mathematical or logical abstractions. Regardless of subject area, modeling and simulation software projects move through similar development iterations: identifying and defining a problem, figuring out what requirements that problem implies, considering what designs might those requirements, model construction, simulation operation and – assuming the software becomes a tool – some kind of maintenance. Also, modeling and simulation projects that intend to impact decision making have stakeholders, who may include sponsors, users, builders (Balci and Sargent 1982). Even computational cognitive science R&D models involve some level of decision support – for example, as when the model us used to make investment decisions about experimental priorities.

Secondly, we emphasize that the Memory and Reasoning project is a computational science R&D project; and as such, a systematic approach to comparing a code's results to empirical data supports the quality of the project. Verification and validation activities help the project characterize and map the evolving functional domain of the constructed software, as well as identifying the formal demands one can and should place on the testing.

Related to the above, verification and validation establish a body of evidence that a code is performing to a set of predetermined specifications. In other words, verification and validation are part of the larger project of computational **model evaluation**. By **model evaluation**, we mean *the systematic assembly of evidence sufficient to provide a comprehensive evaluation of the modeling and simulation software's performance with regard to a set of requirements* (preferably specified before any code gets written). Such an evaluation could include issues of usability and design, as well as the correctness of the model – evaluation aspects that are captured in the terms verification and validation.

Challenges in Applying ASC V&V Approaches to Social and Cognitive Modeling

That said, directly and specifically applying ASC V&V principles to cognitive modeling and simulation is a challenge for several reasons. The first has to do with the nature of R&D software as a particular kind of software project. The ASC approach to verification and validation entails a hierarchical view of the functionality that is encompassed by software, and how that functionality might evolve, and what the understood functionality targets might be. This hierarchical emphasis is captured in many of the documents that elaborate ASC V&V planning specifications, the notion of a Phenomenology Identification and Ranking Table, emphasis on the intended application of the software rather than the software itself, gap analysis to support prioritization of V&V testing in the face of resource constraints, and the notion of a Predictive Capability Maturity Model. While it might be difficult in some cases to develop a crystal clear understanding of a the hierarchy imposed by functionality within implemented software, most software engineering "best practice" reflects the view that software expresses functional capabilities that one should be able to completely, unambiguously, and objectively specify.

Unfortunately, **this last statement is *incorrect* for R&D software!** And this is one major challenge for adapting requirements-oriented verification and validation to software that intends to support the production of new knowledge. In an R&D project, we may not necessarily know the desired functionality, nor do we necessarily know whether the functionality we have, or are aiming for (say in the V&T document) is correct. For example, in implementing an additional element of functionality, we could learn that everything we implemented to that point is wrong. This point raises to the level of a new art form the way software functionality, both achieved and desired, should be specified, and what the most important initial conditions are for performing useful V&V, especially validation.

Related to this issue is the fact that verification and validation is a methodology that uses "referents" to "quantify" in a controlled circumstances the "quality" of a computational model (code) for use in a specific "application." Each of the terms in quotation marks is a charged term: their meaning is both clear and important in the context of V&V for computational modeling and simulation in the physical sciences, but are less so in the areas of computational social and cognitive science. In fact, each of these terms – referent, quantify, quality, and application – highlights an area of research and development for modeling and simulation evaluation in computational social and cognitive science.

Thirdly, the concept of validity in psychological research is a complicated one. Reber's Penguin Dictionary of Psychology includes twenty-eight distinct definitions of validity, each of which has a distinct meaning in relation to experimental psychology. For example, face validity asks whether concepts are appropriate at a first glance; while construct validity evaluates whether or not a test represents the phenomenon it purports to represent; and content validity asks whether the tests chosen to assess the model are appropriate and specific to the model's claims.

In contrast, David Hopkin divides validity into two basic categories: content validity emerges from a qualitative or subjective assessment of a theory, model, or experiment; while criterion validity sets out objective metrics against which a theory, model, or experiment can be compared and assessed. In regards to evaluating modeling and simulation software, Hopkin adds a third element of credibility; namely reliability, which assesses the software's consistency and dependability in performance over time.
Lastly, computational cognitive and social science differ from the physical and mathematical sciences in that theoretical or conceptual models of the phenomena of interest may not enjoy unified support across the epistemic community of interest. In fact, computational cognitive and social modeling and simulation software may actually

be a platform for developing and exploring new theories for multilevel, psycho-social-physiological phenomena that are not clearly understood. This situation stands in stark contrast to the physical sciences: while model form may be a topic of debate, the theoretical principles that describe the phenomena under study are rarely (if ever) an area for debate. In other words, at the theoretical level at least, many of the physical sciences tend to enjoy a level of consensus that the social and cognitive sciences do not. This means that verification and validation activities for computational cognitive and social science must place particular emphasis on the selection and development of the *conceptual model* that is going to be implemented in the code.

In making this point, we are following the guidance offered by the Defense Modeling and Simulation Office (DMSO), which defines a simulation conceptual model as "a bridge between [model] developer and [model] user…." (Defense Modeling and Simulation Office (DMSO) 2006). Conceptual models not only support communication among and between model developers In physics, a conceptual model is "composed of all mathematical modeling data and mathematic equations that describe the physical system or process of interest… [it[is produced by analyzing and observing the physical system…" (Oberkampf, et al. 2003). However, conceptual models may be described qualitatively as well, and represented graphically, mathematically, in narrative, or in a combination of formats. and stakeholders; they also help the team translate the intended application area into a plan for the model, insofar as a clear description of the concepts and relationships helps identify what needs to be in the code, and points the way to possible valiation activities. Moreover, when the theoretical framework for explaining a phenomenon of interest is emergent or young (and subject to debate), a conceptual model identifies authoritative justification for why particular constituent elements are included in the model. We discuss conceptual model development and documentation in greater detail below.

V&V Methodology for Computational Cognitive Modeling and Simulation

In the previous section, we discussed the major concepts that structure the environment in which verification and validation methodologies have been developed and applied in the nuclear weapons program. In this section, we discuss the means by which they can be adapted to cognitive and social modeling. We discuss several levels of activity to contribute to the evaluation of the CS&T software, beginning the high-level assembly and documentation of program goals and project requirements, then moving to specific ideas for planning and executing verification and validation activities. In doing so, we identify three primary components of the V&V methodology that ASC has developed that should be adopted for rigorous V&V of cognitive and social model. These are (1) planning; (2) verification methodology; and (3) validation methodology. In the context of model evaluation, we propose that

- ***Verification*** refers to methods that help determine the internal logical correctness, consistency, sufficiency and accuracy of a computational model; and
- ***Validation*** refers to methods intended to gather evidence of the external logical correctness, consistency, sufficiency, and accuracy of a computational model.

We emphasize that, particularly in the context of R&D modeling and simulation efforts, absolute verification and validation of a "code" is an unwieldy goal, simply because R&D software has a discovery goal that modeling and simulation software derived from stable knowledge often does not. Secondly, we recognize that, as a methodology for accumulating evaluative evidence about the performance of a code, verification and validation can become time consuming, expensive, intimidating, requiring significant investments in documentation and record keeping, in addition to planning and resource investments. Before long, V&V activities can take on the characteristics of an expensive and even annoying obstruction to completing a modeling and simulation project, rather than a supportive evaluation methodology. Given that V&V methods are unlikely to provide a necessary and sufficient basis of evidence for judging the "truth" carried in a code, what is the point of engaging in this activity?

We see several reasons for attempting to systematically evaluate the goodness of R&D computational modeling and simulation projects. As we discuss below in greater detail, verification and validation emphasize a thoughtful, judicious approach to the development of R&D software, beginning with the development of the problem space, the formulation of a theoretical framework, the development of a conceptual model and its instantiation into code, the evaluation of that software, and an assessment of the model's utility as a predictive application. Moreover, as Trucano, Pilch and Oberkampf point out, there are both minimal and maximal levels of validation, depending on the performance requirements of the code project (Trucano, et al. 2002); in other words, verification and validation activities may be tailored to the level of external scrutiny that the code project expects from its stakeholder audience, who may be located in academia, industry, or other national laboratories.

We have tried to be as thorough as possible in describing a maximal set of V&V activities in this document, but that does not mean that the M&R team needs to pursue all the tasks we have identified below. The level of verification and validation that is appropriate for the project is up to the code development team, which needs to consider its goals for the code in light of available resources and stakeholder expectations in order to identify an optimal and workable V&V strategy.

This brings us to the first component of verification and validation: namely, planning. Experience in the ASC program has demonstrated that, in developing a modeling and simulation project that is capable of being verified and validated, nothing is more important than specifying the M&S application precisely and rigorously. This is because, as economist Clive Granger pointed out, the criteria used to evaluate the simulation derive from the stated application of the model (Hendry and Ericcson 2001). In this section, we describe several elements that might be included in a planning process for code projects in computational cognitive and social science. This consists of three kinds of activity, each building on its predecessor: identification of programmatic goals; identification of prokect goals and the derivation of requirements; and identification of an appropriate theoretical approach and documentation of the conceptual model. These three activities collectively form the basis for the Phenomenon Identification and Ranking Table.

As we mentioned earlier, we used Sandia's Cognitive Science and Technology (CST) program, and the Memory and Reasoning(M&R) Project in the CS&T program, as a context in which to consider what V&V might look like when extended outside the ASC program into new disciplinary domains. Arrows (⇒) indicate how the suggested tasks align with elements of the M&R Design Document, which is included in Appendix B.

Identify Programmatic Goals

A first stage of planning requires involves documenting how the modeling project supports the programmatic goals of the Sandia research investment area that supports the project. (⇒ *M&R Design Document, Section 1*). Figure 8, below, outlines suggested tasks and output for early planning discussions.

Figure 10: Tasks and Output for Early Stage Planning Discussions

Given that Sandia is working to develop a mature Cognitive Science and Engineering research program, standards against which we seek to evaluate simulation software should resonate with the goals of the program that is funding and supporting the software development project. Programmatic goals for modeling and simulation can range from the development of insight generating, "a-ha!" tools not intended for use in high-consequence decision making; to predictive models whose outputs will contribute to high-impact decisions (where "impact" describes the ramifications of a decision for a particular constituency, set of stakeholders, or community). For example, the Mission, Vision, and Strategic Goals of the ASC program call for predictive, cutting-edge simulation capabilities supporting decision making for long-term support of the nuclear stockpile. As Trucano et al point out (Trucano, et al. 2005), this implies a pretty high bar for ASC simulations, and means that that "V&V [is]… key to understanding the confidence in these computational tools, and for establishing sufficiency of confidence for the intended applications to the US nuclear weapons program." ASC simulations are instead explicitly aimed at synthesizing knowledge toward key decisions; the guidelines of the ASC V&V program are correspondingly formal, and code evaluation can be quite extensive (and expensive) depending on the intended application of the model.

In contrast, modeling and simulation programs that are developing tools to support academic knowledge production and exploratory thinking are not likely to play a role in official decision-making. Moreover, they will likely be embedded – hopefully deeply so – in ongoing research efforts, and will evolve and grow as research tools in a context of experimentation, theory construction, publication, and peer review. In each context, verification and validation may entail different standards – though as we discuss below, “different” does *not* imply “less rigorous.”

In this stage of planning, specific tasks would include a joint project/management review of any CS&T strategic planning documentation, and a discussion of the CS&T programmatic goals with CS&T and the Memory and Reasoning project investigators and stakeholders. Documentation of these meetings, describing the alignment between the programmatic goals of the CS&T research area and the Memory and Reasoning Project, should be documented in a white paper that points to criteria for model evaluation and assessment.

Identify Project Goals and Derive Requirements

Verification and validation demonstrates that a simulation satisfies specific requirements, which developers derive from an agreed-upon application of the model. This stage of planning focuses on the modeling project itself, and asks the team to consider the goals of their modeling project, the research questions they seek to answer, and the code requirements that these goals and questions imply for the project **(***maps to ⇒ M&R Design Document, Section 1*.) More specifically, this entails specifying an application area; identifying set of questions for which the model will provide outputs; and documenting the resulting requirements for both the model's construction and its performance. Figure 9 illustrates the tasks and outputs for this class of evaluation activities.

In this effort, we want to establish as complete as possible a description of the model intentions, application areas, questions, goals, and the expected user community. In the case of the M&R model, we perceive that the modeling and simulation goals are oriented toward exploring basic research questions. This implies a user community comprising informed subject matter experts who understand the basic phenomena being explored. Planning, then, might entail identifying and documenting central questions to which the researchers are interested in applying the model, then using those questions to identify the specific psychophysiological phenomena that those questions imply. This stage could also involve the prioritization of goals and questions, to differentiate “must answer” questions from others that may be addressable in further iterations of the modeling process. If the team intends that a wider user community adopt its modeling software, then documenting the credibility standards of that community may be useful as well.

Specific tasks at this stage include a review of the documentation about the M&R model to identify application areas, specific question sets, and expected outputs; discussions with researchers and model developers to prioritize question sets and goals; documentation of these discussions; and review with stakeholders.

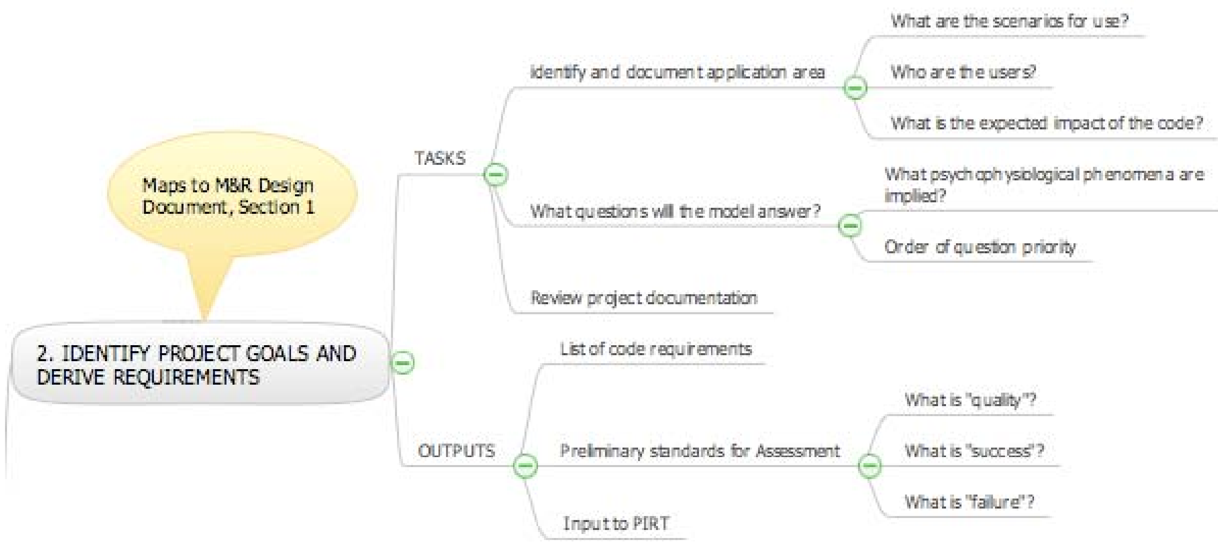


Figure 11: Document Project Goals and Requirements: Tasks and Outputs

Outputs could include a set of documented software/code requirements, preliminary standards for assessment (including definitions of “quality,” “success” and “failure”) and a preliminary list of inputs for a Phenomenon Identification and Ranking Chart (PIRT). Figure 9, above, displays tasks, including suggested questions; and outputs for this stage of planning.

Identify Project Goals and Derive Requirements

This set of activities focuses on the systematic justification and documentation of basic theory, its applicability to the phenomena of interest, and its instantiation as model-in-code **(***maps to ⇒ M&R Design Document, Sections 2 and 3*). It includes a review of the theoretical approaches to the problem of interest; selection of the “best” approach(es) to inform model development; formalize the theoretical approach into a conceptual model; and to create a basis for transparency in describing the process through which the conceptual model is transformed into a computationally tractable entity and instantiated in software (Activity 4).

The major output from this stage of the project planning and documentation is the Phenomenon Identification and Ranking Table. This is a key planning methodology and representational format that identifies and prioritizes future activities, but it also formalizes the project’s thinking about the problem at hand. As such, we have identified a series of activities that building an effective computational social or cognitive science R&D PIRT requires: Identifying a range of theoretical approaches to the problem, justifying a theoretical approach, formalizing a conceptual model, developing documentation, and organizing peer review sessions. We discuss and illustrate each of these areas of activity separately below.

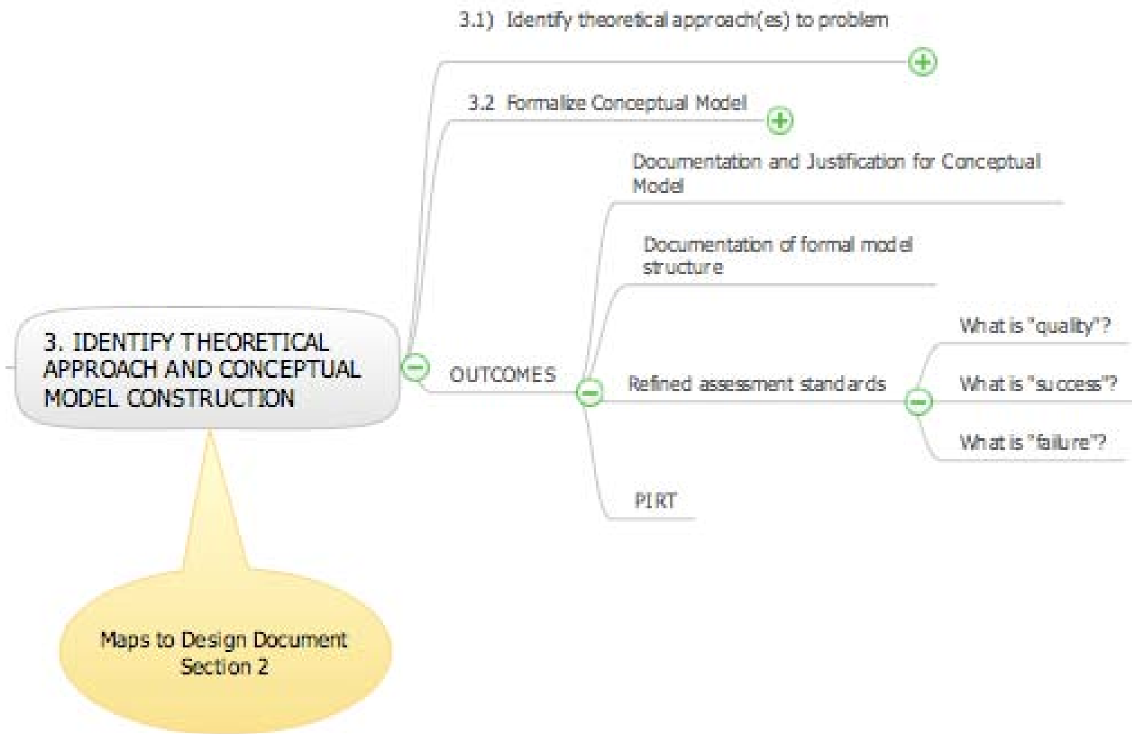


Figure 12: Theoretical Approach and Conceptual Model

Identifying Theoretical Approaches to the Problem

Identifying a theoretical approach to the problem requires that the team work to iteratively between the problem space and literature and existing research that helps the team capture the current state-of-knowledge in the wider research community about the problem at hand (*maps to ⇒ M&R Design Document, Sections 2.1 and 2.2*). In identifying the project goals and the derivation of requirements, the modeling team has specified at least the preliminary psychophenomenological phenomena that are of research interest (this may have occurred during the proposal process, or in a preliminary planning document). Identifying the theoretical approaches to the problem involves a more thorough elicitation and documentation of the main categories of phenomena that comprise the problem space; for example, the processes and sub-processes that are being modeled and the relationships among them. This specification will likely emerge as the team reviews the literature and develops a theoretical basis for the approach that it is identifying for addressing the problem space. This likely entails an extensive literature review and documentation of the main constructs that others have used to address the area of interest; consultation with outside experts; and an effort to identify gaps in what is known and possible formulations that address those gaps.

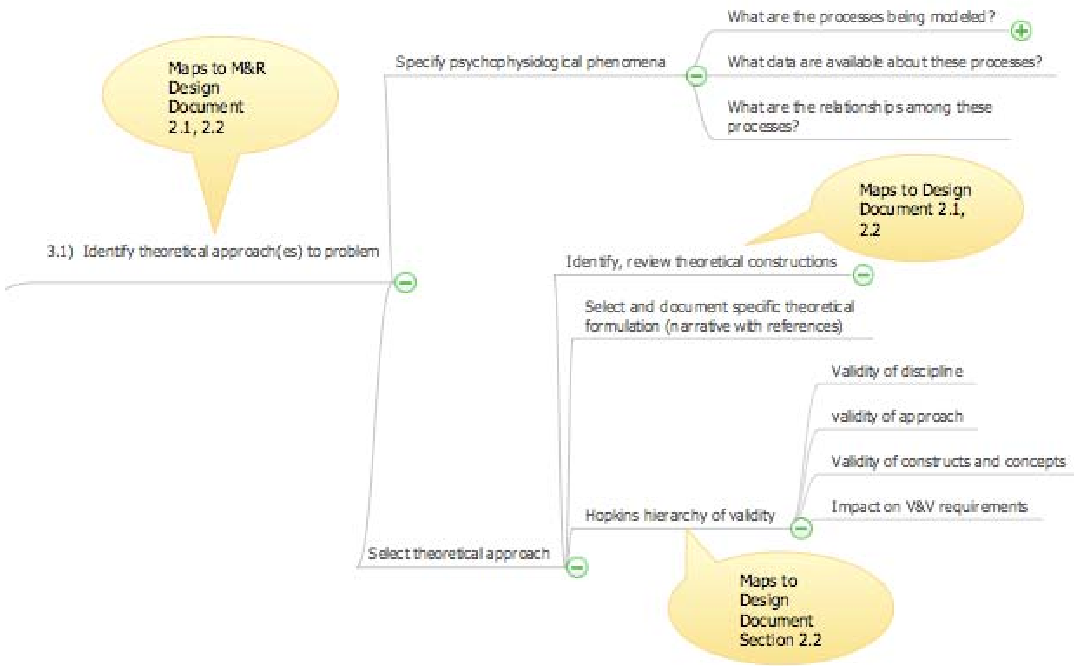


Figure 13: Identifying, Documenting, and Considering Validity of Theoretical Approach

The Hierarchy of Conceptual Validity: David Hopkin

In this regard, we believe that David Hopkin's discussion of validity as a hierarchical problem is an important rubric for assessing not only the state of knowledge about a problem space, but the level of acceptance that the modeling and simulation team is likely to meet as it disseminates its approach into the relevant epistemic community (Wise and Hopkin 1992). Hopkin identified three levels at which the problem of validity for conceptual models needs to be considered, the first of which is the validity of the discipline itself: is the discipline recognized as a legitimate form of inquiry, with methods, theories, datasets, and practitioners engaged in the progressive production of reliable knowledge about reality? Secondly, within the discipline, is the approach that the team is taking, theoretically and methodologically, recognized as an appropriate frame for the problem at hand? Lastly, what is the recognized validity of the constructs and concepts that the team is using to address the problem: to what extent can the team uncover evidence of consensus about how to define, implement, measure, and bound the concepts that are used to describe and/or explain the problem at hand?

Figure 14: Hopkin's Hierarchy of Validity (from Wise and Hopkin 1992)

These three types of validity – discipline, approach, and constructs/concepts – create a *hierarchy of validity* that has implications for many aspects of the modeling and simulation project. For example, if the disciplinary context is not considered valid among the stakeholder community (e.g., using astrological theories to explain mental illness to an audience of psychologists) then the modeling team can expect the project to encounter a great deal of critique. Dealing with that critique will require extensive documentation and accumulation of evidence to support the team's approach to framing and operationalizing the problem space it is working in. Moreover, if the discipline is not recognized, then neither the approach nor the selected constructs nor concepts will be viewed as valid. Even if a disciplinary context is widely accepted in the scientific community as legitimate, within that context, a particular approach to a problem might be contentious (e.g., explaining mental illness as a product of mainly environmental and not physiological factors). If the approach is not perceived as legitimate, then the constructs and concepts entailed in the selected approach are also unlikely to enjoy legitimacy. The team will have to thoroughly document and justify why it has selected a more controverisal framing and how it is encoding that framing in the simulation, as well as the evidence that will be used to evaluate the approach. Lastly, the team may be pursuing a valid approach to a valid problem that is recognized by the members of a legitimate disciplinary community (modeling the interaction of environmental and physiological factors in the onset of clinical depression). If the phenomenon is not well or adequately explained by existing constructs or concepts entailed within that approach – and in most research environments, the development of demonstrably adequate explanatory constructs is what makes research exciting – then the modeling project can expect to have to document and justify as transparently as possible how it has framed the problem, and the implications of that framing for methodological decisions (using computational modeling and simulation) and data sources. Considering Hopkin's hieararchy of validity in relation to the problem space, the approach, and the core concepts and methods will support the development of the PIRT.

Formalize Conceptual Model

Guidance from the Defense Modeling and Simulation Office (DMSO) represents some of the most articulate thinking about the role of a conceptual model in translating the goals of a modeling and simulation process into specifications for what will go in the software, when, and how (*maps to ⇒ M&R Design Document, Sections 2.3, 2.4, 2.5, 2.6 and Section 3*). As such, the conceptual model is one of the most important elements of a modeling and simulation project because it serves as a bridging abstraction that brings key information from the previous planning discussions into the software development process. In a nutshell, a conceptual model takes all the requirements identified in the previous discussions and lays them out as specifications for software design and implementation.

In a research environment, a conceptual model specifies **the hypotheses** the model intends to address, so that the software includes all the entities that are necessary to run the simulation and generate valid data. It describes the computational and organizational context in which the model is expected to run, the kinds of decisions it will likely support, and - particularly in a research environment – the role of the model in supporting the production of knowledge about a particular class of research problems. The conceptual model identifies key elements, entities and processes; defines the level of accuracy and precision that is required of each element; describes the relationships among them, and identifies sources of data and information to characterize them in the model. It defines the model's architecture, the algorithms that will be used (or developed), and the model, It lists and characterizes input and output variables and explains what the model is doing with inputs to generate outputs. Lastly, the conceptual model helps identify validation referents, or "the best [empirical] information available that describes characteristics and behavior of the reality represented in the system" (Defense Modeling and Simulation Office (DMSO) 2006). In other words, the conceptual model identifies early on the empirical, measurable characteristics of the real-world process or event that the model intends to capture. As we have noted above, the issue of "referents" may be a real problem in computational cognitive and social science. We discuss referents in greater detail below.

Software Implementation

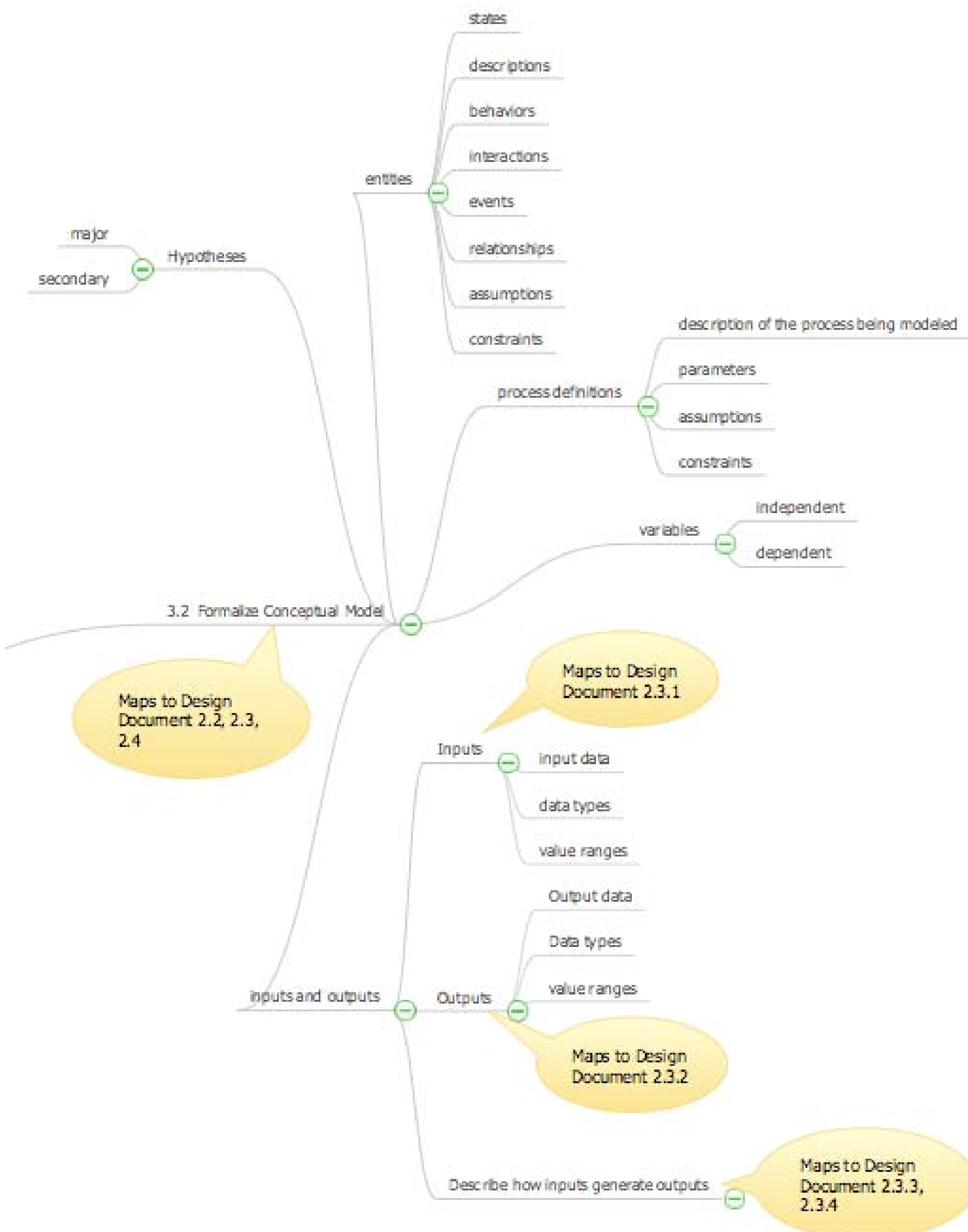
DMSO has identified six principles for translating a problem (or "mission space") into a set of statements to guide the development of a simulation. This guidance emphasizes necessity, sufficiency, parsimony, justification, and transparency. The principles include the following ideas (Defense Modeling and Simulation Office (DMSO) 2006):

- Ensure that every simulation element is identified, described, and represented in the conceptual model.
- Ensure that anything the model is going to assess is represented by a specific simulation element, clearly identified and described in the conceptual model
- Every simulation element should have a "real world" counterpart. Remember that data and metadata structures can have a huge impact on the design, specification and implementation of a simulation element.
- Whenever possible, the simulation elements should correspond to accepted paradigms so that the model is more understandable, transparent, and acceptable to outsiders. If you are developing a new algorithm or paradigm, document it thoroughly, including its relationship to/deviation from other algorithms or paradigms.
- If some component of the simulation doesn't meet these principles, but you consider it necessary, include it; but document it and use it only when absolutely critical.
- Don't include anything that doesn't need to be in the simulation.

One way of ensuring the completeness of the conceptual model is to ask if a developer unfamiliar with the project could take the conceptual model and use it as a map to implement the ideas in software.

Conceptual models are related to validation in two ways: validating a conceptual model is important in generating information related to the overall credibility of software simulation; and the conceptual model supports simulation validation, as it helps to identify and document the elements that represent the phenomena and sub-phenomena being studied.

- Validating a conceptual model requires a peer or external reviewer assessment process in which a group of knowledgeable individuals who are familiar with the problem domain, and with the process of model construction review the conceptual model for scope, completeness, level of detail, and the accuracy with which it represents the phenomenon of interest. This is something of a challenge for research software, as the mechanism that generates the phenomenon of interest might itself be the object of the research process. In R&D software, then, a conceptual model review might assess the quality of the hypotheses, their theoretical justification (as supported by existing literature and data), and the adequacy with which the proposed implementation will enable the researchers to explore the phenomenon of interest.
- The conceptual model also supports verification and validation activities. The process of assessing (validating) the conceptual model should be documented as a source of information for demonstrating the credibility of the model. The conceptual model itself provides validation guidance: it identifies referents that



states
descriptions
behaviors
interactions
entities
events
major
Hypotheses
relationships
secondary
assumptions
constraints
description of the process being modeled
parameters
process definitions
assumptions
constraints
independent
variables
dependent
3.2 Formalize Conceptual Model
Maps to Design Document 2.3.1
Maps to Design Document 2.2, 2.3, 2.4
Inputs
input data
data types
value ranges
Output data
Data types
value ranges
inputs and outputs
Outputs
Maps to Design Document 2.3.2
Maps to Design Document 2.3.3, 2.3.4
Describe how inputs generate outputs

Figure 15: Formalizing the Conceptual Model

Key Output: The PIRT

As we have discussed earlier, one key challenge for verification and validation of computational cognitive and social science modeling and simulation applications is the presence of disagreement among or within disciplines about the descriptive or explanatory frameworks (theories) that apply to a phenomenon of interest. We believe that difficulties in conceptual model validation are prominent in research-centric model development, where the goal of the modeling and simulation effort is not to instantiate a well-understood model in code, but to create a code that will help define a theoretical basis for describing and/or explaining a phenomenon of interest (such as the formation and retention of memory in the human brain). This is an important theme in both computational cognitive and social modeling projects (see for example (Ball 2007; Epstein 2006; Epstein 1999).

Here we acknowledge a significant point of divergence between the physical and social sciences. In the former, there is a relatively broader base of consensus about the validity, applicability, and mathematical specification of theoretically-derived models. In the social sciences (which include such disciplines as economics, political science, sociology, psychology, anthropology, even history) phenomena of interest may be described by multiple, even competing theoretical paradigms, each of which likely derive from different disciplines, and which can marshal various sources of empirical evidence to support their claims. For example, a cultural anthropologist studying race and inequality might use ethnographic data to develop a behavioral model that describes the discursive maintenance of racial difference; while a psychological anthropologist might look to biological explanations for the way perceive and act upon physical difference; while an economist might draw from decision theory to explain how individual choice patterns maintain conditions of inequality (and all three are as likely to borrow from each other, as they are to identify inadequacies in the others' conceptualizations). Social phenomena are extremely complex, often emergent, and therefore difficult to study experimentally; hence consensus about the appropriateness of underlying theory and its specification in a model cannot be taken for granted. This is particularly true in cutting-edge areas of social science research, or in sites of inquiry where paradigmatic shifts are underway; or in very complex, multilevel problems where the researcher is seeking to couple individual behaviors to broad social phenomena. In fact, computational modeling and simulation projects often explicitly seek to generate new explanatory frameworks, rather than instantiate existing models derived from generally accepted theoretical premises (Epstein 2006; Epstein 1999).

The absence of a well-understood and accepted theoretical foundation for the developed conceptual model is problematic for systematic efforts at experimental validation in cognitive/social models. The question then becomes how one can appropriately design and implement V&V activities to support a somewhat research-centric model development effort that, nonetheless, has a long-term goal of providing "mature" application relevance. This is one of the research themes in our own work. It is typical to simply dismiss V&V because of this concern, but that is not a useful view in our opinion.

We believe that transparency in the selection of disciplinary and theoretical approaches; the selection of a specific conceptual model; the process through which that model is abstracted and formalized into operational concepts and relationships; and the instantiation of that model into a computationally tractable algorithm are critically important of external stakeholders are to be able to appreciate and evaluate the model *on its own terms*.

We see the PIRT as an important tool in developing a documentation trail to support efforts to address this problem. As discussed above, the Phenomenology Identification and Ranking Table (PIRT) is a core planning methodology in ASC Program, and drives validation activities. In the context of computational cognitive and social science models, the PIRT can play an important role not only in code validation, but in early-stage conceptual model validation as well, insofar as it helps bound and decompose the complexity of the conceptual model issues.

For one thing, our experience in the ASC program has demonstrated that more controversial the PIRT, the less likely is validation. This is because a lack of consensus about the important phenomena, their importance relative to the application area, and the best means to evaluate their external correspondence to the real world indicates significant epistemic uncertainty about either the phenomenon that is being modeled, the modeling approach, the software itself, or all three. Validation under conditions of controversy and uncertainty is difficult.

Extending this principle to computational cognitive and social science, developing a PIRT provides multiple benefits in a computational cognitive or social science research and development project, where controversy over the selection of a theoretical framework and its instantiation in code is likely. The PIRT forces the modeling and simulation R&D team to clearly identify and articulate the constituent phenomena and to rank them in a hierarchy of relevance; to identify the current state of validity for the conceptual explanations of the phenomena; and to specify how the conceptual phenomena are captured in the code and what experiments are most important in accumulating evidence that the conceptual model captured in the code is the correct one. R&D teams that cannot agree on what should go into the PIRT or how the elements should be ranked are likely dealing with a problem space characterized by a high degree of epistemic (knowledge) uncertainty. When the basic constitutive elements of a broader phenomenon are not well-enough understood to be specified and ranked, it is difficult to systematically plan a set of experimental validation and/or data gathering activities. In assessing the level of importance and validity of different elements of the conceptual model, it may be useful to conduct a formal elicitation session using a comparative methodology such as the Analytic Hierarchy Process (AHP).

As a planning tool, the PIRT also supports decisions about the best investment of limited R&D resources. A very important phenomenon that is poorly understood may demand a relatively high level of validation attention; conversely, a less-critical phenomenon that is well-characterized, which broadly acceptable model forms, probably requires fewer validation resources.

Verification and Validation Activities for Computational Cognitive and Social Science

In this section, we discuss activities five and six, Verification and Validation, as a joint core effort through which a majority of evaluation-class data will be generated. At some level, verification and validation is all about specifying, constructing and executing good tests. Of particular importance is the documentation of the testing rationale, design, execution, and results, including any error bars or other indicators of uncertainty. Accordingly, we spend some time discussing testing, then touch briefly on verification and validation methodologies specific to computational cognitive science.

Some Words on Testing

Testing means applying a test engineering methodology to the software in question. This includes elements like the following, all of which should be clearly and explicitly addressed, and documented:

Specification. What is the test is and why was it chosen? This specification should involve an understanding of what software is being touched, that is *covered*, by the test. The point is that the test is providing V&V evidence only for the software that it is covering. The purpose of the specification is so that the world can precisely understand the definition of the test and why the project considers it to be useful. This allows the project to to repeat the test at will – and allows skeptics in the rest of the world to do so as well.

Coverage. Clearly, defining test coverage is very important, but unfortunately the software engineering literature – which is enormous – is inconclusive in defining what coverage actually entails. There are several metrics that get used; perhaps the most common metric of coverage is line coverage. This is problematic when although even the definition of a software line is ambiguous, as it can be in certain software languages). Regression testing is often coupled to metrics of line coverage, as in "What percentage of the code is covered by the regression test suite." This question is not necessarily sensible for verification tests, however, as regression tests are not automatically verification tests. This is because regression testing is a software engineering technique, not a V&V technique. That said, verification tests that are run "in regression" with the software development effort are certainly "regression tests."

The most important coverage may actually involve functionality; this is the tack that the ASC program has taken. Coverage and functionality are somewhat tricky in a research context, as functionality can shift and change depending on the evolution of the hypotheses and theoretical framing. If V&V aims to provide functionality coverage for the Memory and Reasoning software, then verification and validation needs to responds software life cycle; that is, to the version evolution. At the same time, V&V also will influence the version evolution as the comparisons between simulation outputs and referents shape the thinking of the team about the problem. This relationship between evaluation and software development – which is so rich and complex in the context of a software research effort – justified the amount of work that goes into explaining what the test is doing, why it was chosen, what the pass/fail criteria are, and what referents are being used. This brings us to the topic of referents, which we touched on briefly in the discussion of conceptual models above.

Referents: Both verification and validation depend on referents; that is, a baseline specification determined to be the "correct" answer or behavior for the test. A referent can be specified many ways, but if a referent cannot be specified, then verification and validation are impossible. Specification of the "referent" is a requirement for a test to support verification and validation.

- *Verification Referents*: Referents for verification are often a source of confusion. Verification referents are internal referents; they are characterized as mathematical or as software metrics. Verification referents *cannot* be specified as an experimental observation, because observed behavior in, say, people, has no relevance whatsoever to whether or not software bugs exist, or mathematical errors underlie a software implementation. *Internal tests require internal referents*.

- *Validation Referents*: In contrast, a validation referent is an external referent that is used to assess the relationship between the simulation's results and the real-world phenomenon of interest. This means that validation requires an experimental observation as a referent. Mathematical theorems and software engineering principles say nothing about whether the underlying functionality that, in the M&R software, attempts to model human memory.

- *Summary*: Nothing is more important than understanding that "external" referents do not provide a means of assessing whether functionality has been implemented correctly, and that "internal" referents do not say anything whether the implemented functionality is correct.

The issue, or properly stated, the *challenge of determining valid referents,* especially for validation, is paramount. R&D projects pose heavy burdens for determining referents. After all, if we already knew what was "correct," why do we have to do research? In the context of research, a referent should probably be related to the hypotheses that are being tested, a source of data for experimental comparison to support or diminish the acceptability of the hypothesis which the model is implementing. Hypothesis testing, modeling and simulation, and the problem of validation referents are deeply intertwines.

*Implementation and Execution***.** How is the test implemented? Some tests may need to be executed repetitively and systematically for each new version of the software – in which case the test iself should be under version control. Tests may be implemented as part of a test suite that the entire project owns; or may be implemented by an individual contributor to the project. Similarly, information about test execution (schedule, ownership, results) requires documentation. This information may become quite elaborate if the software is being tested for multiple computing platforms, as is the norm in ASC.

Analysis deals with the results of executing a test. Analysis is about the information from the executed test required for comparison to the identified referent. The work required to extract this information needs to be specified as clearly as everything else. In a perfect world, all tests, including validation tests, are so simple that this element does not introduce complexity. This is the real world, so expect that extracting comparison-quality information will be more time consuming that expected.

Comparison involves putting the test results next to the referent and assessing the delta between the two. We sometimes use the word "metric" to imply this comparison. It is important to remember that uncertainty that needs to be quantified to properly compare the analysis to the referent. If the computational model fully emulates the observed referent, then the comparison step requires dealing with uncertainty. In psychology, this is always specified using stochastic methods (statistical procedures). This means that statistical procedures need to be used to compare model and observation. This is suddenly a nontrivial, not necessarily well-posed problem. On the other hand, the comparison could be as direct as the relative examination of two numbers. Unfortunately, even in this case there is still the lurking problem of deciding what the comparison means in terms of "correctness" of the functionality and its implementation.

Pass/Fail. This is the final level of assessment: the decision about whether or not the model passed or failed the test, based on the comparison of the referent to the analysis results, and taking uncertainty into account. Ideally, pass-fail criteria are established prior to execution of the test. This requires some polarization in thinking, and that can be challenging in the context of an evolving software project.

Implication. What is the final conclusion to be drawn from performing the test, the analysis and the Pass/Fail assessment? This is especially important in open life cycle models, and in computational science R&D software projects, where the results of testing can have immediate and devastating impact on one's faith in implemented functionality and one's plans for future functionality. A relatively few very crucial validation tests, properly executed and properly evaluated, could have implications that destroy years of work. Verification and validation are a stern test of the software development project.

Now that we have discussed testing, we turn to verification and validation.

Verification

Verification is the accumulation of evidence that desired model ***functionality*** is correctly implemented. Implementation, in this case, means "software implementation." Software implementations rest on mathematically correct algorithms, which in turn rest on mathematically correct theories or broader formulations. In computational science and engineering, verification centers on the correct or accurate transformation of an underlying conceptual model into a computational model through the creation of a mathematical formalism; development of numerical algorithms for solving the mathematical equations, and implementation of the numerical algorithms in software ("code").

This process is rather straightforward in fields such as computational physics. The mathematical underpinnings of most computational physical science conceptual models are often very straightforward and so conventional that there is little argument that certain equations, such as the equations of continuum mechanics, essentially define the conceptual models underlying thermal, fluid and solid mechanics. Accordingly, verification activities are relatively straightforward as well (if time consuming) and typically involve some combination of both code verification and solution verification, depending on the maturity of the code and previous verification activities. Oberkampf et al (Oberkampf and Trucano 2002) describe a set of verification activities that should accompany the creation of the model as it moves from mathematical formalism to algorithms to code implementation. To the degree that verification can be seen as more than testing is because of the mathematical content that is present in computational science.

Verification is more problematic for cognitive/social models. If as assume that evaluation of the *model translation process* – from real world problem to conceptual model to code – is the focus of verification activities, then verification in the cognitive and social science modeling and simulation world must emphasize the selection and documentation of the conceptual model, the abstractions and representations that are used to capture it (mathematical, logical, graphical, narrative), and the translation of those abstractions into some computationally tractable form. In this realm, the role of mathematics and mathematical formulations is far more problematic: for one thing, the essence of the mathematical/logical core of a cognitive or social code may be not be a transparent set of (Oberkampf and Trucano 2002)equations; the core might instead be a set of logical rules, which become the explicit targets for the algorithms implemented in the associated code. If one could prove that software implementations were correct then one would not need to perform verification testing (the goal of Formal Methods). Because Formal Methods approaches are not yet applicable in the realm of computational cognitive modeling and simulation, well-chosen, structured and executed tests will provide evaluation data for judging implementation for the time being.

That said, we do believe that the logical skeleton suggested above, in which verification examines the correctness of the results of passing from mathematics to algorithms to software, is adaptable to cognitive and social computational modeling and simulation projects. In all computational science R&D projects, regardless of discipline, principles of good software engineering and code testing do apply. Software engineering tests, such as regression tests, are not verification tests, but they do help catch bugs in the code as it evolves and are important in assessing the overall quality of the code.

Specific verification tests should be designed based on the mathematics that express the conceptual model, and which are in turn captured in algorithms and software. For example, verifying an Adaptive Resonance Theory (ART) network implementation may include evidence that mathematical theorems about the characteristics of ART networks have been proved. An ART network requires a broad mathematical conception that should be correct; it is reduced to a mathematical algorithm, or collection of algorithms; and those algorithms are implemented in specific software languages on specific computers. Verification of ART requires accumulating evidence that all of this has been correctly performed.

The M&R project has adopted some of the principles of verification described above and has developed documents to identify specific tests that might support verification. For example, their V&V document describes several possible verification tests for an operational M&R computational model. The document describes several types of verification tests, but – following our comments about the importance of the relationship between V&V activities and evolving functionality – what is lacking is testing strategy that is dovetailed with the model version evolution. We believe that this is more a matter of documentation and getting details under control than it is of knowledge and expect the project to continue evolving its testing plans as part of an overall testing strategy.

Validation

Validation can be thought of as completely defined testing, and requires comparing computational results with observations gathered from a clearly defined experimental setting. Validation is completely dependent on two things: identifying a *validation referent*, or a known point of truth for comparison that enables one to evaluate the correctness of the model; and the ability to generate real-world data (preferably experimentally) around that referent.

These requirements for referents and empirical data create something of a conundrum for research codes that intend to generate data in areas where epistemic (lack of knowledge) uncertainty predominates. This is because generating referents requires some idea of what constitutes a ground truth about the phenomenon of interest; and in a research environment, the whole point of the modeling and simulation effort may very well be aimed at establishing a better explanation for what is going on.

The problem of validation is particularly difficult in the social sciences, particularly those that deal with group behavior (sociology, economics, anthropology). We are aware of many modeling and simulation efforts in the social sciences that aim at generating theories or explanations for poorly understood phenomena (Ball 2007; Epstein 2006; Epstein 1999). Modeling and simulation is a methodology that is attractive because it enables the researcher to simulate something that resists experimentation or even empirical observation – for example, about the movement of ideas through large populations, or the emergence of intergroup violence.

In the social sciences, there are basic debates about the role of theory as a descriptive, explanatory, or causal framework; and whether or not a nomothetic enterprise is even possible (i.e., the generation of broadly applicable, generalizable explanatory theories for human behavior). As Jessica Turnley points out, evaluation techniques that rest on a logical positivist philosophy that assumes the existence of objective data and which presumes stable relationships between data and theory are a poor fit for the social sciences, where multiple frameworks can be evoked with equal credibility, depending on one's discipline, to explain similar phenomena. As a result, the social sciences, the problem of how to interpret data – that is, how to establish reliable, stable frames of explanation that are supported by empirical evidence. Moreover, some forms of social knowledge resist quantification, or may lose their value when quantified – as when documenting the nuances of how people adopt a belief system are reduced to counting the number of individuals who choose to attend a particular church (Turnley 2004). And last, but certainly not least, validation quality data might demand conducting an experiment – something that is difficult to do when the phenomenon of interest involves the dissemination of an idea through a large population, or assessing the causes of intergroup violence in a particular region of the world.

In regards to validation, these problems create quite a challenge for assessing the "truth" of a model's results. We take a pragmatic approach to these problems. As we have argued above, we are interested in leveraging the ASC work in verification and validation to formulate a strategy for evaluating computational modeling and simulation projects in other disciplines. Models may not be validatable (Konikow and Bredehoeft 1992); but that does not mean that *evaluation* is impossible. Indeed, we have argued that verification and validation really represent the systematic generation, documentation and accumulation of evidence about the performance of a model in relation to a given application area.

If we think of validation more broadly as a matter of getting the best possible evidence to assess the fit between the model and the phenomenon of interest, then referent has to be the firmest basis of evidence possible that can be used to serve as a point of comparison between a model's predictive output and the real-world phenomenon. Obtaining validation quality data entails identifying the right sources of data; and gathering those data correctly – in other words, doing the right experiment, and doing that experiment right. What "right data" and "right collection" entail will depend on the standards of the discipline whose members are the audience for the modeling project's results.

Model Assessment

We still need to worry about the way comparison is made. In strict verification and validation, this requires the establishment of a *validation metric*, a pass/fail mark for evaluating the difference between the model results and the referent. In modeling and simulation applications that rely on partial differential equations (PDEs) as the mathematical representation of the problem space, this is a relatively straightforward problem, as PDEs generate output that is quite easily compared to experimental results in engineering and physics experiments. Hence, if we let

M=Model, R=Referent and e=a threshold for pass/fail, then

$||M-R||<e$?

represents a clear, quantitative metric. This is a yes/no question, and in the physical sciences, where experimental criteria are clear, we are likely dealing with powerful referents that are unlikely to be disputed as validation quality data.

However, as we move from PDEs to computational cognitive and social science, we lose metric strength, as the ability to compare outputs to empirical evidence is less straightforward. Assessing the difference between M and R and figuring out what that difference means may be a matter of judgment on the part of people who are interested in applying the model in a particular decision space. A limiting case of this kind of assessment is some complex perception of model utility with no explicit metric of any kind: for example, U(M,R) in which U is some decisionmaker's level of "happiness" with the model. The decision maker acknowledges the existence of R, but does not consider it a test, and black-and-white questions about passing or failing a test are meaningless.

Uncertainty Quantification

This brings us to the topic of uncertainty, which plays a critical role in the evaluation of the model to the referent The greater complexity of dealing with uncertainty in cognitive and social models is probably the single greatest challenge in validating cognitive and social models, particularly in comparison to physical models. Again, we turn to the computational science and engineering literature as a source of ideas for thinking about the problem of uncertainty in the cognitive and social sciences.

The literature on uncertainty quantification recognizes two kinds of uncertainty. Aleatory uncertainty points to random variability and is typically quantified using statistical methodologies. Epistemic uncertainty, on the other hand, points to a lack of knowledge about the phenomenon at hand. Probability is used to describe epistemic uncertainty, as when Bayesian subjective assessments that draw on expert judgment are used to develop a statistical model; or when a distribution is put on a distribution to characterize the uncertainty around the true value of a poorly understood parameter. Epistemic uncertainty is also characterized by so-called Generalized Information Theories (Klir 2003) based on possibility and probability theory. Alternatives include set-valued probabilities (fuzzy probability, for example) that can incorporate "soft" uncertainty characterizations.

IN CLOSING: SOME COMMENTS ON VERIFICATION AND VALIDATION FOR MODELS IN COMPLEX DECISION ENVIRONMENTS

As we have discussed above, computational modeling and simulation software is increasingly being developed for use in complex policy environments where technical, social, economic, political and even psychological phenomena come into play. Verification and validation become particularly challenging in this area, particularly in comparison to computational science and engineering. For example, in computational science and engineering, determining whether a software tool is accurately solving a set of partial differential equations (verification) is a logically internalized process. It requires no engagement with the world of observation; it requires no experimentation. Similarly, assessing whether or not an agent-based model accurately executes a conceptual model that defines the movements of a terrorist cell requires the ability to rigorously assess the mathematics, algorithms, and software engineering of the system.

On the other hand, determining whether a partial differential equation is correct does require engagement with the external world. Correctness is not determined by mathematical logic, but must be centered on observations derived from controlled experiments. In contrast, assessing whether the agent-based simulator is built on correct requirements, one of which has to be something like "accurately predicts one or more features of terrorist-cell movement," requires comparison with observation. The logic of math, algorithms, and software engineering alone are simply not adequate for validation of this kind of simulation; the world external to the simulator is required. Performing a systematic and meaningful comparison between the simulation and the external world requires referents, which are critical for "validating" computational tools for our most complex decision support needs. Just as in computational science and engineering, referents require empirical observations. However, the quality of referent in computational science and engineering is likely quite different than that which might be available in policy realms, where controlled experimentation is impossible.

Reinforcement of Application Emphasis

All realistic V&V centers on a clear and precise specification of the intended area for the modeling and simulation application. V&V creates a domain of credibility for the modeling and simulation application that targets that area. The execution and results of verification and validation are strongly constrained by the application. The precision required to understand an application to achieve V&V will have broader utility for research projects in complex research and development environments that support high consequence decisions, and must therefore balance a wide set of participants, goals, and methodologies.

Risk-Informed Decision Analysis (RIDA)

No important decision that involves risk is *based* on the results of a computational tool. Rather, a modeling and simulation application informs decisions. That is, modeling and simulation applications are simply one of many elements that enter into the decision making environment. The model itself may be (likely is) a contributor to the risk influencing the decision making environment. We certainly expect that the decision making environments in national security will involve high risk. Therefore, we see a need for decision support tools that function properly within *risk-informed decision analysis* (RIDA). Our past and current practice and experience with verification and validation increasingly emphasizes the importance of model evaluation role in assessing the credibility of software application components for decision-support within risk informed decision analysis. To the extent this is possible, V&V is then a foundational element that clarifies the risks inherent in the decision, as well as assesses the credibility of modeling and simulation application decision support tools.

Uncertainty Quantification (UQ)

Verification and validation are a critical organizing principle and quantitative framework for performing UQ. There is no verification and validation without uncertainty quantification. In this role, verification and validation likely lead to a broader understanding and use of UQ within an overall decision making environment. This is another cross-cutting role for evaluation research in modeling and simulation research programs.

Model-Observation Integration

Validation requires observation-based referents (external logic-based referents). Validation thus integrates computational tools and the experiments-observations (external referents) that are required to assess their validity. Because this modeling and simulation application-referent integration is a requirement for validation, it provides significant opportunities and mechanisms for enriching the interaction between observations (external referents) and modeling and simulation application across an entire project. This is a third cross-cutting role for verification and validation.

Credibility Specification

Systematic model evaluation is virtually the only objective basis for rigorously defining the credibility of modeling and simulation application tools. Verification and validation contribute to the rigor of mode-supported decision through the methodologies they offer for assessing of modeling and simulation application credibility.

Computational Tool Management

A critical issue in model development for decision making is the development of instruments that allow cost-benefit analysis, prioritization, and gap management for modeling and simulation application decision-support tools. Much of this information is created by rigorously executed verification and validation, and can be organized to support needs like cost-benefit analyses. A current example of a complex metric that organizes information for these purposes is the ***Predictive Capability Maturity Model (PCMM)*** used in the ASC program, which supports additional conceptual integration between decision environment and model.

From Information to Knowledge

One common view of complex decision making environments is that they must facilitate the passage of ***information to knowledge to wisdom (I2K2W)***. Informally speaking, if we regards "the dots" as information, then knowledge means "connecting the dots," and wisdom means using the connected dots wisely in decision making. Computation in and of itself does not typically provide knowledge, nor does it necessarily facilitate the conversion of a pool of information into knowledge. V&V is a necessary condition for achieving knowledge from computational information processing or creation. I could further argue that V&V is therefore relevant to creating the ability to optimize decisions based on the existing knowledge base, in other words, to make “wise” decisions. It is therefore imperative to understand modeling and simulation application decision-support tools within this conceptual structure.

Constraining Expectations

Neither knowledge nor wisdom is an automatic product of computational decision support tools because of limitations of verification and validation and other factors. Certainly the constraints on modeling and simulation application credibility that rigorous verification and validation provides play an important role in constraining expectations associated with modeling and simulation application decision support tools. This is certainly one way to avoid the tendency, if there is one, to “throw technology over the fence” and hope it is used wisely.

Managing Organizational Factors

We have noticed that modeling and simulation application decision support tools can actually create organizational (that is, decision environment) stress and are therefore often a disruptive, rather than helpful, technology. One of our working hypotheses is that verification and validation processes can serve to to create a formal Community of Practice, or a collaborative circle of combined stakeholders, whose work can enable healthier organizational engagements with modeling and simulation application decision support tools. Even from our experience gained with the ASC program, we have seen that V&V contributes to better organizational understanding of the strengths and weaknesses of modeling and simulation application decision support tools. Such organizational understanding is imperative for managing the I2K2W passage. V&V in this sense is part of the glue that integrates the decision environment. We see verification and validation as an important methodology helps transform disruptive problems in the construction, implementation, and application of modeling and simulation application decision support tools, perhaps by simply contributing to healthy recognition of creative tensions among modelers, stakeholders, and decision makers.

REFERENCES

Adrion, W. Richards, Martha A. Branstad, and John C. Cherniavsky
1982 Validation, Verification, and Testing of Computer Software Computing Surveys 14(2).
Agar, M. H., et al.
2004 Epidemiology or Marketing? The Paradigm Busting Use of Complexity and Ethnography. *In* Proceedings of Agent 2004: Challenges in Social Simulation.
American Institute of Aeronautics and Astronautics
1998 Guide for the Verification and Validation of Computational Fluid Dynamics Simulations (AIAA G-077-1998). Reston, VA: American Institute of Aeronautics and Astronautics.
Ang, James A., Timothy Trucano, and David R. Luginbuhl
1998 Confidence in ASCI Scientific Simulations. Pp. 22. Albuquerque, NM: Sandia National Laboratories.
Axelrod, Robert
2003 Advancing the art of simulation in the social sciences University of Michigan.
Axtell, Robert, et al.
1996 Aligning Simulation Models: A case study and results. Computational and Mathematical Organization Theory 1(2):123-141.
Balci, O.
2003 Verification, Validation, and Certification of Modeling and Simulation Applications Simulation Conference 2003.
Balci, Osman, and R. G. Sargent
1982 Validation of Multivariate Response Trace-Driven Simulation Models
Ball, Philip
2007 Social Science Goes Virtual. Nature 448(9 August 2007):647-648.
Bankes, Steve
1993 Exploratory Modeling for Policy Analysis. Operations Research 41(3):435-449.
Bernard, Michael L.
2008 Design Document for the Neurobiology of Recollection: A Computational Model. Pp. 40. Albuquerque, NM: Sandia National Laboratories.
Card, Stuart K., Jock Mackinlay, and Ben Shneiderman
1999 Information Visualization. *In* Readings in Information Visualization: Using Vision to Think. S. K. Card, J. Mackinlay, and B. Shneiderman, eds. San Francisco, CA: Morgan Kaufmann.
Carley, Kathleen M.
1996 Validating Computational Models: Carnegie Mellon University.
Christensen, Randy
1998 ASCI Three Dimensional Hydrodynamic Instability and Turbulence Modeling. *In* Science and Technology Review. Livermore, CA: Lawrence Livermore National Laboratory.
Cornwall, John M.
2005 Preliminary Report of the ASC Predictive Science Panel.
Davis, Paul K.

1992 Generalizing Concepts and Methods of Verification, Validation, and Accreditation (VV&A) for Military Simulations: RAND.

Defense Modeling and Simulation Office (DMSO)

2006 Conceptual Model Development and Validation: RPG Special Topic. Washington, DC: Department of Defense, Defense Modeling and Simulation Office.

Enthoven, Alain C., and K. Wayne Smith

1971 How Much is Enough? Shaping the Defense Program, 1969-1969. New York: Harper and Row.

Epstein, Joshua

2006 Generative Social Science: Studies in Agent-Based Computational Modeling. Princeton, NJ: Princeton University Press.

Epstein, Joshua M.

1999 Agent-Based Computational Models and Generative Social Science. John Wiley & Sons, Inc. 4(5).

Fagiolo, Giorgio, Christopher Birchenhall, and Paul Windrum

2007 Empirical Validation in Agent-Based Models. Computational Economics 30(3).

Fagiolo, Giorgio, Paul Windrum, and Alessio Moneta

A Critical Guide to Empirical Validation of Agent-Based Economics Models: Methodologies, Procedures, and Open Problems.

Fagiolo, Giorgio, Paul Windrum, and Alessio Moneta

2006 Empirical Validation of Agent-Based Models: A Critical Survey: Laboratory of Economics and Management Sant'Anna School of Advanced Studies.

Forrester, Jay W.

1971 Counterintuitive Behavior of Social Systems Theory and Decision 2:109-140.

Forrester, Jay W., and Peter M. Senge

1980 Tests for Building Confidence in System Dynamics Models TIMS Studies in the Management Sciences 14:209-228.

Futral, W. Scott, et al.

1999 Performance of ALE3D on the ASCI Machines. *In* Nuclear Explosives Code Development Conference. Pp. 6. Las Vegas, Nevada: Lawrence Livermore National Laboratory.

Goldstein, Harry

2006 Modeling Terrorists: New Simulators Could Help intelligence Analysts Think Like the Enemy. IEEE Spectrum (September 2006):26-35.

Hendry, David F., and Neil R. Ericcson

2001 Understanding Economic Forecasts. Cambridge, MA: MIT Press.

Hodges, Ann, et al.

2001 ASCI Software Quality Engineering: Goals, Principles and Guidelines. Pp. 12 with 2 appendices. Albuquerque, NM: Sandia National Laboratories.

Kerbyson, Darren J., Adolfy Hoisie, and Harvey J. Wasserman
2005 A Performance Comparison between the Earth Simulator and Other Terascale Systems on a Characteristic ASCI Workload. Concurrency and Computation: Practice and Experience 17:1219-1238.
Klein, Richard, et al.
2006 ASC Predictive Science Academic Alliance Program Verification and Validation White Paper. Livermore, CA: Lawrence Livermore National Laboratory.
Klir, George
2003 An Update on Generalized Information Theory. *In* International Symposium on Imprecise Probabilities and Their Applications. J. M. Bernard, T. Seidenfeld, and M. Zaffalon, eds. Lugano, Switzerland: Carleton Scientific
Konikow, Leonard F., and John D. Bredehoeft
1992 Ground-water Models Cannot be Validated. Advances in Water Resources 15:75-83
Kusnezov, Dimitri F.
2004 Advanced Simulation and Computing: The Next Ten Years, Office of Advanced Simulation and Computing, NNSA Defense Programs, issued by Sandia National Laboratories, SAND 2004-3740P.
Logan, Roger, and Cynthia Nitta
2002 Verification and Validation Methodology and Quantitative Reliability at Confidence: Basis for an Investment Strategy. Pp. 109. Livermore, CA: Lawrence Livermore National Laboratory.
Mackenzie, Donald
1995 The Automation of Proof: A Historical and Sociological Exploration. IEEE Annals of the History of Computing 17(3):7-29.
Mackenzie, Donald
2001 Mechanizing Proof: Computing, Risk and Trust. Cambridge, MA: MIT Press.
Menzies, Tim
2002 Verification and Validation in Artificial Intelligence. Foundations 02: A V&V Workshop, Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland, USA, 2002, pp. 70.
Menzies, Tim , and Charles Pecheur
2005 Verification and Validation and Artificial Intelligence. Advances in Computing 65:154-203.
Na Alessa, Lilian, Melinda Laituri, and Michael Barton
2006 An "All Hands" Call to the Social Science Community: Establishing a Community Framework for Complexity Modeling Using Agent Based Models and Cyberinfrastructure. *In* Journal of Artificial Societies and Social Simulation, Vol. 9.
Nitta, Cynthia, and Roger Logan
2004 Qualitative and Quantitative Linkages from V&V to Adequacy, Certification, and Benefit/Cost Ratio. Pp. 87. Livermore, CA: Lawrence Livermore National Laboratory.

Nowak, David A., and Ronald A. Christensen
1997 ASCI Applications. *In* Supercomputing 1997. Pp. 12. San Jose, CA: Lawrence Livermore National Laboratory.
O'Keefe, Robert M., and Daniel O'Leary
1993 Expert System Verification and Validation: A Survey and Tutorial Artificial Intelligence Review 7:3-42.
Oberkampf, William L., Timothy G. Trucano, and Charles Hirsch
2004 Verification, validation, and predictive capability in computational engineering and physics. Applied Mechanics Reviews 57(5):345-384.
Oberkampf, William L., Timothy G. Trucano, and Charles Hirsch
2002 Verification and Validation in Computational Fluid Dynamics in Progress in Aerospace Sciences, Volume 38, 209-272 (2002)
Oberkampf, William L., Martin Pilch, and Timothy G. Trucano
2007 Predictive Capability Maturity Model for Computational Modeling and Simulation, SAND2007-5948. Albuquerque, NM: Sandia National Laboratories.
Pilch, Martin, et al.
2001 Guidelines for Sandia ASCI Verification and Validation Plans - Content and Format: Version 2.0. Pp. 82. Albuquerque, NM: Sandia National Laboratories.
Pilch, Martin, et al.
2004 Concepts for Stockpile Computing (OUO Report). Pp. 109. Albuquerque, New Mexico: Sandia National Laboratories.
Polhill, Gary J., Luis R. Izquierdo, and Nicholas M. Gotts
2005 The ghost in the model (and other effects of floating point arithmetic). Journal of Artificial Societies and Social Simulation 8(1).
Post, Douglass E, and Richard P Kendall
2004 Software Project Management and Quality Engineering: Practices for Complex, Coupled Multiphysics, Massively Parallel Computational Simulations: Lessons Learned from ASCI. The International Journal of High Performance Computing Applications 18(4):17.
Roache, Patrick J.
1998 Verification and Validation in Computational Science and Engineering. Albuquerque, NM: Hermosa Publishing.
Sargent, Robert G.
1979 Validation of simulation models. Winter Simulation Conference, San Diego, CA, United States, 1979, pp. 497-503. IEEE Press.
Sargent, Robert G.
1985 An Expository on Verification and Valifidation of Simulation Models. Winter Simulation Conference, 1985, pp. 15-22.
Sargent, Robert G.
1987 An Overview of Verification and Validation of Simulation Models. Winter Simulation Conference, 1987, pp. 33-39.

Skias, Spiro T.
2006 Background of the Verification and Validation of Neural Networks. *In* Methods and Procedures for the Verification and Validation of Artificial Neural Networks. B. J. Taylor, ed. Pp. 1-12 New York: Springer.
Trucano, Timothy
2005 Uncertainty in Verification and Validation: Recent Perspectives. *In* SIAM 2005 CS&E Meeting. Pp. 45. Albuquerque, NM: Sandia National Laboratories.
Trucano, Timothy, et al.
2001 Description of the Sandia Validation Metrics Project: Sandia National Laboratories.
Trucano, Timothy, Martin Pilch, and William L. Oberkampf
2002 General Concepts of Experimental Validation of ASCI Code Applications. Pp. 137. Albuquerque, NM: Sandia National Laboratories.
Trucano, Timothy, et al.
2005 Software Engineering Intersections with Verification and Validation (V&V) of High Performance Computational Science Software: Some Observations. Pp. 15. Albuquerque, NM: Sandia National Laboratories.
Turing, Alan
1950 Computing Machinery and Intelligence. Mind (236):433-460.
Turnley, Jessica G.
2004 Validation Issues in Computational Social Simulation. Pp. 8: Galisteo Consulting Group.
Wilenski, Uri, and William Rand
2007 Making models match: Replicating an agent-based model. Journal of Artificial Societies and Social Simulation 10 (42).
Windrum, Paul, Giorgio Fagiolo, and Alessio Moneta.
2007 Empirical Validation of Agent-Based Models: Alternatives and Prospects. Journal of Aritificial Societies and Social Simulation 10(2).

APPENDIX A: Working Bibliography

Ackoff, R. L.

1988/89 From Data to Wisdom: Presidential Address to ISGSR, June 1988. Journal of Applied Systems Analysis 16(1998):7.

Adamic, Lada A., and Eytan Adar

2005 How to Search a Social Network. Social Networks 27:187-2003.

Adrion, W. Richards, Martha A. Branstad, and John C. Cherniavsky

1982 Validation, Verification, and Testing of Computer Software Computing Surveys 14(2).

Agar, M.H., et al.

2004 Epidemiology or Marketing? The Paradigm Busting Use of Complexity and Ethnography. *In* Proceedings of Agent 2004: Challenges in Social Simulation.

Aharoni, Amikam

1995 Agreement between theory and experiment. Physics Today 48(6):33-37.

Aigner, Dennis J.

1972 A Note on Verification of Computer Simulation Models. Management Science 18(11):615-619.

Alan, A., and B. Pritsker

1997 Modeling in Performance-Enhancing Processes. Operations Research 45(6):797-804.

Alberti, Marco, et al.

2004 The SOCS Computational Logic Approach to the Specification and Verification of Agent Societies. *In* Lecture Notes in Computer Science (LNCS). C. Priami and P. Quaglia, eds. Pp. 314-339, Vol. 3267/2005. Berlin: Springer.

Alfarano, Simone, Friedrich Wagner, and Thomas Lux

2004 Estimation of Agent-Based Models: The Case of an Asymmetric Herding Model

Allen, Nicholas A., et al.

2003 Next generation modeling II - applications: improving the development process for eukaryotic cell cycle models with a modeling support environment. New Orleans, Louisiana: Winter Simulation Conference.

Allen, Nicholas A., Clifford A. Shaffer, and Layne T. Watson

2005 Building modeling tools that support verification, validation, and testing for the domain expert. Orlando, Florida: Winter Simulation Conference.

Allen, P. Geoffrey, and Bernard J. Morzuch

2006 Twenty-Five Years of Progress, Problems, and Conflicting Evidence in Econometric Forecasting. What About the Next 25 Years?. International Journal of Forecasting 22:475-492.

Alqallaf, Fatemah, and Paul Gustafson

2001 On Cross-Validation of Bayesian Models. The Canadian Journal of Statistics 29(2):333.

Althaus, C. E.

2005 A Disciplinary Perspective on the Epistemological Status of Risk. Risk Analysis 25(3):567-588.

Amar, Robert, and John Stasko

2004 A Knowledge Task-Based Framework for Design and Evaluation of

Information Visualizations. *In* IEEE Symposium on Information Visualization. Pp. 143-149. Austin, TX: IEEE.

American Institute of Aeronautics and Astronautics
1998 Guide for the Verification and Validation of Computational Fluid Dynamics Simulations (AIAA G-077-1998). Reston, VA: American Institute of Aeronautics and Astronautics.

Ammerman, Albert J.
1992 Taking Stock of Quantitative Archaeology. Annual Review of Anthropology 21:231-249.

Anagnostopoulous, Dimosthenis, and Mara Nikolaidou
2006 Data Organization and Data Comparison for Model Validation in Faster-than-Real-Time Simulation. Proceedings of the 39th annual Symposium on Simulation.

Anastasio, Michael, and David Cooper
1998 The Accelerated Strategic Computing Initiative: A Daunting Challenge. *In* Science and Technology Review, Vol. April 1998: Lawrence Livermore National Laboratory.

Andrus, D. Calvin
2005 The Wiki and the Blog: Toward A Complex Adaptive Intelligence Community. Pp. 28. Washington, DC: Central Intelligence Agency.

Ang, James A., Timothy Trucano, and David R. Luginbuhl
1998 Confidence in ASCI Scientific Simulations. Pp. 22. Albuquerque, NM: Sandia National Laboratories.

Apostolakis, George E.
2004 How useful is quantitative risk assessment? Risk Analysis 24(3).

Armstrong, J. Scott
2007 Significance Tests Harm Progress in Forecasting. Journal of Forecasting 23:321-327.

Armstrong, Scott J.
2006 Findings from Evidence-Based Forecasting: Methods for Reducing Forecast Error International Journal of Forecasting 22:583-598.

Armstrong, Scott J.
2007 Statistical Significance Tests are Unnecessary Even When Properly Done and Properly Interpreted: Reply to Commentaries. Journal of Forecasting 23:335-336.

Arthur, James D., et al.
1999 Evaluating the Effectiveness of Independent Verification and Validation. Computer 32(10):79-83.

Arthur, James D., and Richard E. Nance
2000 V\&A; III: verification and validation without independence: a recipe for failure. Orlando, Florida: Society for Computer Simulation International.

Asad, Ch. Ali, Muhammad Irfan Ullah, and Muhammad Jaffar-Ur Rehman
2004 An Approach for Software Reliability Model Selection. COMPSAC'04, 2004.

Axelrod, Robert
2003 Advancing the art of simulation in the social sciences University of Michigan.

Axtell, Robert, et al.
1996 Aligning Simulation Models: A case study and results. Computational and

Mathematical Organization Theory 1(2):123-141.
Baines, Tim, et al.
2004 Humans: The Missing Link in Manufacturing Simulation? Simulation Modelling Practice and Theory 12(7-8):515-526.
Balci, Osman
1986 Credibility Assessment of Simulation Results. Winter Simulation Conference, 1986, pp. 38-44
Balci, Osman
1988 The implementation of four conceptual frameworks for simulation modeling in high-level languages. Winter Simulation Conference, San Diego, California, United States, 1988, pp. 287-295. ACM.
Balci, Osman
1988 The implementation of four conceptual frameworks for simulation modeling in high-level languages. http://doi.acm.org/10.1145/318123.318204 Winter Simulation Conference, San Diego, California, United States 1988 pp. 287-295 ACM.
Balci, Osman
1989 How to Assess the Acceptability and Credibility of Simulation Results. Winter Simulation Conference, 1989, pp. 62-71.
Balci, Osman
1990 Guidelines for successful simulation studies (tutorial session) Winter Simulation Conference, New Orleans, Louisiana, United States, 1990, pp. 25-32 IEEE Press.
Balci, Osman
1994 Validation, verification, and testing techniques throughout the life cycle of a simulation study Winter Simulation Conference Orlando, Florida, United States, 1994, pp. 215-220 Society for Computer Simulation International.
Balci, Osman
1995 Principles and Techniques of Simulation Validation, Verification and Testing. Winter Simulation Conference, 1995, pp. 147-154.
Balci, Osman
1997 Verification validation and accreditation of simulation models. Winter Simulation Conference, Atlanta, Georgia, United States, 1997, pp. 135-141. IEEE Computer Society.
Balci, Osman
1998 Verification, validation, and accreditation. Winter Simulation Conference, Washington, D.C., United States, 1998, pp. 41-4. IEEE Computer Society Press.

Balci, Osman
2001 A methodology for certification of modeling and simulation applications ACM Trans. Model. Comput. Simul. 11 (4):352-377
Balci, Osman
2003 Verification, Validation, and Certification of Modeling and Simulation Applications Simulation Conference 2003.
Balci, Osman, et al.
2002 A Collaborative Evaluation Environment for Credibility Assessment of Modeling and Simulation Applications Winter Simulation Conference 2002.
Balci, Osman, and Richard E. Nance
1987 Simulation support: prototyping the automation-based paradigm Winter Simulation Conference, Atlanta, Georgia, United States 1987 pp. 495-502 ACM.
Balci, Osman, et al.
2002 Expanding our horizons in verification, validation, and accreditation research and practice. Winter Simulation Conference, 2002. Vol. 1, pp. 653-663.
Balci, Osman, et al.
2002 Improving the model development process: expanding our horizons in verification, validation, and accreditation research and practice. San Diego, California: Winter Simulation Conference.
Balci, Osman, et al.
1990 Model generation issues in a simulation support environment *In* Winter Simulation Conference Pp. 257-263 New Orleans, Louisiana, United States IEEE Press.
Balci, Osman, and William F. Ormsby
2000 Well-defined intended uses: an explicit requirement for accreditation of modeling and simulation applications. Winter Simulation Conference, 2000, pp. 849-854.
Balci, Osman, et al.
2000 Planning for Verification, Validation and Accreditation of Modeling and Simulation Applications. Winter Simulation Conference, 2000, pp. 829-839.
Balci, Osman, and Said D. Saadi
2002 Simulation standards: proposed standard processes for certification of modeling and simulation applications Winter Simulation Conference, San Diego, California, 2002, pp. 1621-1627, Winter Simulation Conference.
Balci, Osman, and Robert G. Sargent
1981 A Methodology for Cost-Risk Analysis in the Statistical Validation of Simulation Models. Communications of the ACM:190-197.
Balci, Osman
1982 Some Examples of Simulation Model Validation Using Hypothesis Testing. SESSION:621-629.
Balci, Osman, and R. G. Sargent
1982 Validation of Multivariate Response Trace-Driven Simulation Models
Ball, Philip
2007 Social Science Goes Virtual. Nature 448(9 August 2007):647-648.

Bankes, Steve
1993 Exploratory Modeling for Policy Analysis. Operations Research 41(3):435-449.
Barlas, Yaman
1989 Multiple Tests for Validation of System Dynamics Type of Simulation Models European Journal of Operational Research 42:59-87.
Barlas, Yaman
1990 An autocorrelation function test for output validation. Simulation
Barlas, Yaman
1996 Formal Aspects of Model Validity and Validation in System Dynamics System Dynamics Review 12(3):183-210.
Barreteau, Olivier
2003 Our companion modeling approach. Journal of Artificial Societies and Social Simulation 6(1).
Beck, M. B.
1987 Water quality modeling: A review of the analysis of uncertainty. Water Resources Research 23(8):1393-1442.
Becker, Joerg, Bjoern Niehaves, and Karsten Klose
2005 A framework for epistemological perspectives on simulation. Journal of Artificial Societies and Social Simulation 8(4).
Bederson, Benjanim B., and Ben Shneiderman
2002 The Craft of Information Visualization: Readings and Reflections. San Francisco: Morgan Kaufmann.
Beizer, Boris
1984 Software system testing and quality assurance. New York: Van Nostrand Reinhold Co.
Beizer, Boris
1990 Software Testing Techniques New York Van Nostrand Reinhold.
Bell, Peter C., and Robert M. O'Keefe
1995 An Experimental Investigation into the Efficacy of Visual Interactive Simulation. Management Science 41(6):1018-1038.
Ben-Haim, Yakov
2004 Uncertainty, probability and information-gaps. Alternative Representations of Epistemic Uncertainty 85(1-3):249-266.
Bendell, Tony
1986 Minimising Misconceived Models. The Statistician 35(3):303-309.
Bernard, Michael L
2008 Design Document for the Neurobiology of Recollection: A Computational Model. Pp. 40. Albuquerque, NM: Sandia National Laboratories.
Besag, Julian, et al.
1995 Bayesian Computation and Stochastic Systems. Statistical Science 10(1):3-41.
Beusmans, Jack, and Karen Wieckert
1989 Computing, Research, and War: If Knowledge is Power, Where is Responsibility? Communications of the ACM 32(8):9.

Beven, Keith
2002 Towards a Coherent Philosophy for Modeling the Environment. Proceedings: Mathematical, Physical and Engineering Sciences 458(2026):2465-2484.
Bianchi, Carlo , et al.
2005 Validation and calibration in ACE models: An investigation of the CATS Model Universita di Parma, Italy.
Bonabeau, Eric
2002 Agent-Based modeling: Methods and Techniques for Simulating Human Systems. PNAS 99(3):7280-7287.
Booker, Jane, and Laura McNamara
2005 Expert Knowledge in Reliability Characterization: A Rigorous Approach to Eliciting, Documenting, and Analyzing Expert Knowledge. *In* Engineering Design Reliability Handbook. E. Nikolaidis, D. M. Ghiocel, and S. Singhal, eds. Pp. 13-1 - 13-31. Boca Raton, FL: CRC Press.
Booker, Jane M., et al.
2004 An Engineering Perspective on UQ for Validation, Reliability and Certification. Pp. 30. Los Alamos, NM: Los Alamos National Laboratory.
Boswijk, H. Peter , Cars H. Hommes, and Sebastiano Manzan
2005 Behavioral Heterogeneity in Stock Prices University of Amsterdam.
Boudreau, Marie-Claude, David Gefen, and Detmar W. Straub
2001 Validation in Information Systems Research: A State-of-the-Art Assessment. MIS Quarterly 25(1):1-16.
Brade, Dirk
2000 Enhancing modeling and simulation accreditation by structuring Verification and Validation results. Winter Simulation Conference, Orlando, Florida, 2000, pp. 840-848. Society for Computer Simulation International.
Brita, Louis G., and F. Nur Ozmizrak
1996 A Knowledge-Based Approach for the Validation of Simulation Models: The Foundation. ACM Transactions on Modeling and Computer Simulation (TOMACS) 6(1):76-98.
Brown, Daniel, et al.
2005 Path Dependence and the Validation of Agent-Based Spatial Models of Land Use International Journal of Geographical Information Science 19(2):153-174.
Brown, Judy, et al.
1999 Human-Centered Computing, Online Communities, and Virtual Environments. IEEE Computer Graphics and Applications:5.
Brown, Warren B.
1967 Model-Building and Organizations. The Academy of Management Journal 10(2):169-178.
Brun, Xavier F., and Shreyes N. Melkote
2006 Modeling and experimental verification of partial slip for multiple frictional contact problems. 265(1-2):34-41.

Bryant, Randal E.
1991 A Methodology for Hardware Verification Based on Logic Simulation. Journal of the ACM (JACM) 38(2):299-328.
Bryson, Joanna J., Yasushi Ando, and Hagen Lehmann
2007 Agent-based modelling as a scientific method: a case study analysing primate social behaviour Biology 362(1485):1685-1698.
Buckley, Chris and Slaton, Gerard
1995 Optimization of Relevance Feedback Weights. SIGIR '95:7.
Burton, Michael L
1973 Mathematical Anthropology. Annual Review of Anthropology 2:189-199.
Burton, Richard M
2003 Computational Laboratories for Organization Science: Questions, Validity, and Docking. Computational and Mathematical Organization Theory 9:91-108.
Burton, Richard M., and Borge Obel
1995 The Validity of Computational Models in Organization Science: From Model Realism to Purpose of the Model. Computational and Mathematical Organization Theory 1(1):57-71.
Butts, Carter, and Kathleen Carley
2002 Structural Change and Homeostasis in Organizations: A Decision-Theoretic Approach.
Butz, William P., and Barbara Boyle Torrey
2006 Some Frontiers in Social Science. Science 312.
Card, Stuart K., Jock D. Mackinlay, and Ben Shneiderman
1999 Readings in Information Visualization: Using Vision to Think. San Francisco, CA: Morgan Kaufmann.
Carley, Kathleen
1992 Organizational Learning and Personnel Turnover. Organization Science 3(1):20-46.
Carley, Kathleen
2000 Information Security: The Human Perspective.
Carley, Kathleen
2003 Destabilizing Terrorist Networks. Proceedings of the 8th International Command and Control Research and Technology Symposium, National Defense War College, Washington DC, 2003.
Carley, Kathleen, et al.
2004 BioWar: Scalable Multi-Agent Social and Epidemiological Simulation of Bioterrorism Events. NAACSOS Conference 2003, Pittsburgh, PA, 2004. Vol. Day 4.
Carley, Kathleen, and Craig Schreiber
2002 Information Technology and Knowledge Distribution in C3I teams. Proceedings of the 2002 Command and Control Research and Technology Symposium, Naval Postgraduate School, Monterey, CA, 2002.
Carley, Kathleen M.
1996 Validating Computational Models: Carnegie Mellon University.

Carmines, E. G., and R. A. Zeller
1979 Reliability and Validity Assessment. *In* Sage University Paper Series on Quantitative Applications in the Social Sciences, Vol. 07-017.

Carnap, Rudolf
1936 Testability and Meaning. Philosophy of Science 3(4):419-471.

Carr, John T., III, and Osman Balci
2000 Verification and validation of object-oriented artifacts throughout the simulation model development life cycle. Winter Simulation Conference, Orlando, Florida, 2000, pp. 866-871. Society for Computer Simulation International.

Carson, John S., II
2002 Verification validation: model verification and validation. San Diego, California: Winter Simulation Conference.

Cartwright, Nancy
1983 How the Laws of Physics Lie. New York: Oxford University Press.

Castillo, E., C. Solares, and P. Gomez
1998 Tail uncertainty analysis in complex systems. Artificial Intelligence 96(2):395-419.

Cataldo, Marcelo, Kathleen Carley, and L. Argote
2001 The Effect of Personnel Selection Schemes on Knowledge Transfer.

Caughlin, Don
1995 Verification, Validation, and Accreditation (VV&A) of Models and Simulations Through Reduced Order Metamodels. 27th Winter Simulation Conference, Arlington, VA, 1995, pp. 1405-1412.

Caughlin, Don
2000 An integrated approach to verification, validation, and accreditation of models and simulations. 32nd Winter Simulation Conference, Orlando, Florida, 2000, pp. 872-881. Society for Computer Simulation International.

Chen, Li-Chiou, and Kathleen Carley
2001 A Computational Model of Computer Virus Propagation. CASOS Conference 2001, Pittsburgh, PA, 2001. Vol. Day 1.

Chen, Li-Chiou, and Kathleen Carley
2002 Modeling Distributed Denial of Service Attacks and Defenses.

Chen, Li-Chiou, and Kathleen Carley
2003 The Impact of Network Topology on the Spread of Anti-Virus Countermeasures. NAACSOS Conference 2003, 2003. Vol. Day 2.

Chen, Mary I., and Timothy G. Trucano
2002 ALEGRA Validation Studies for Regular, Mach, and Double Mach Shock Reflection in Gas Dynamics. United States.

Chen, Wei, et al.
2004 Model Validation via Uncertainty Propagation and Data Transformations AIAA 42(7).

Chew, Jennifer, and Cindy Sullivan
2000 Verification, Validation and Accreditation in the Life Cycle of Models and Simulations. Winter Simulation Conference, Orlando, Florida, 2000, pp. 813-818. Society for Computer Simulation International.

Christensen, Randy
1998 ASCI Three Dimensional Hydrodynamic Instability and Turbulence Modeling. *In* Science and Technology Review. Livermore, CA: Lawrence

Livermore National Laboratory.
Christodoulou, K., and K. Vlahos
2000 Variable Structure Modeling of Dynamic Industry Systems. The Journal of the Operational Research Society 51(9):1029-1040.
Chwif, Leonardo, et al.
2006 A prescriptive technique for V&V of simulation models when no real-life data are available. Winter Simulation Conference, 2006, pp. 911-918.
Cirillo, Pasquale, Mauro Gallegati, and Pietro A. Vagliasindi
2005 Validation and Calibration in ACE Models: An Investigation on the CATS model.
Clegg, Stewart
2006 Why is Organization Theory so Ignorant? The Neglect of Total Institutions. Journal of Management Inquiry 15:426-429.
Cohen, Michael D., James G. March, and Johan P. Olsen
1972 A Garbage Can Model of Organizational Choice. Administrative Science Quarterly 17(1):1-25.
Conway, R. W., B. M. Johnson, and W. L. Maxwell
1959 Some Problems of Digital Systems Simulation. Management Science 6(1):92-110.
Conwell, Candace L., Rosemary Enright, and Marcia A. Stutzman
2000 Capability maturity models support of modeling and simulation verification, validation, and accreditation. Winter Simulation Conference, Orlando, Florida, 2000, pp. 819-828. Society for Computer Simulation International.
Cooke, Roger
1991 Experts in Uncertainty. Oxford: Oxford University Press.
Corman, Steven R.
2006 Using Activity Focus Networks to Pressure Terrorist Organizations. Computational and Mathematical Organization Theory 12:35-49.
Corno, Fulvio, Matteo Sonza Reorda, and Giovanni Squillero
2004 Evolutionary Simulation-Based Validation. International Journal on Artificial Intelligence Tools 13(4):897-916.
Cornwall, John M
2004 Preliminary Report of the ASC Predictive Science Committee. Los Alamos, NM.
Coughlin, Don
2000 An Integrated Approach to Verification, Validation, and Accreditation of Models and Simulations. *In* 2000 Winter Simulation Conference. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds. Miami, FL.
Coyle, Geoff, and Exelby
2000 The Validation of Commercial System Dynamics Models System Dynamics Review 16(1):27-41.

Craft, Brock, and Paul Cairns
2005 Beyond Guidelines: What Can We Learn from the Visual Information Seeking Mantra? *In* Ninth International Conference on Information Visualization: IEEE.
Crandall, Jacob W., and M. L. Cummings
2007 Developing Performance Metrics for the Supervisory Control of Multiple Robots. HRI'07, Arlington, VA, 2007, pp. 8. ACM.
Dahan Dalmedico, Amy
2001 History and Epistemology of Models: Meteorology (1946-1963) as a Case Study. Archives of the History of the Exact Sciences 55(2001):395-422.
David, Nuno, et al.
2004 The Structure and Logic of Interdisciplinary Research in Agent-Based Social Simulation. *In* Journal of Artificial Societies and Social Simulation, Vol. 7.
David, Nuno, Jaime Simao Sichman, and Helder Coelho
2005 The logic of the method of agent-based simulation in the social sciences: Empirical and intentional adequacy of computer programs journal of Artificial Societies and Social Simulation 8(4).
Davidson, Elizabeth
1997 Changing Frames or Framing Change? Social Cognitive Implications of Organizational Change during IT Adoption. *In* Proceedings of the Thirtieth Hawaii international Conference on System Sciences. Maui, Hawaii: IEEE.
Davis, Paul K.
1992 Generalizing Concepts and Methods of Verification, Validation, and Accreditation (VV&A) for Military Simulations: RAND.
de Quiroz, Jose Eustaquio R., and Maria de Futima Q.V. Turnell
1998 Evaluating the Quality of Human Computer Interfaces According to Specific Contexts. IEEE Transactions (??):1296-1301.
Decoursey, Donn G.
1992 Developing Models with More Detail: Do More Algorithms Give More Truth? Weed Technology 6(3):709-715.
Dee, Dick P.
1995 A Pragmatic Approach to Model Validation Coastal and Estuarine Studies 47.
Defense Modeling and Simulation Office (DMSO)
2006 Conceptual Model Development and Validation: RPG Special Topic. Washington, DC: Department of Defense, Defense Modeling and Simulation Office.
Dekker, D., Kathleen Carley, and David Krackhardt
2002 How Do Social Networks Affect Organizational Knowledge Utilization? CASOS Conference 2002, Pittsburgh, PA, 2002. Vol. Day 3.
Desel, Jorg
2000 Teaching System Modeling, Simulation and Validation Winter Simulation Conference, 2000.
Deutch, John
2005 A Nuclear Posture for Today. Foreign Affairs 84(1):49-60.

Dickinson, David L., and Sean P. A. Drummond
2008 The effects of total sleep deprivation on Bayesian updating. Judgment and Decision Making 3(2):181-190.
DOE
1995 The Stockpile Stewardship and Management Program: Maintaining Confidence in the Safety and Reliability of the Enduring US Nuclear Weapon Stockpile. O.D.P. United States Department of Energy, ed. Pp. 19. Washington, DC.
Dogan, Gokhan
2007 Bootstrapping for Confidence Interval Estimation and Hypothesis Testing for Parameters of System Dynamics Models. System Dynamics Review 23(4):415-436.
Dombroski, Matthew, Paul Fischbeck, and Kathleen Carley
2003 Estimating the Shape of Covert Networks. Proceedings of the 8th International Command and Control Research and Technology Symposium, National Defense War College, Washington DC, 2003.
Doukidis, Georgis I., and Ray J. Paul
1985 Research into Expert Systems to Aid Simulation Model Formulation. The Journal of the Operational Research Society 36(4):319-325.
Dourish, Paul
2006 Implications for Design. *In* CHI 2006: Design: Creative and Historical Perspectives. Pp. 541-550. Montreal, Quebec, Canada: ACM.
Dowling, Deborah
1999 Experimenting on Theories. Science in Context 12(2):261-273.
Downey, Laura L.
2007 Group Usability Testing: Evolution in Usability Techniques. Journal of Usability Studies 2(3):133-144.
Duffy, John
2006 Agent-based models and human-subject experiments. Handbook of Computational Economics 2: Agent-based computational economics
Eardley, D., and JASONs
2005 Quantification of Margins and Uncertainties. D. Eardley, ed. Pp. 35. Washington, DC: National Nuclear Security Administration, Department of Energy.
Easterbrook, Steve, and John Callahan
1998 Formal Methods for Verification and Validation of partial specifications: A Case Study. Journal of Systems and Software 40(3):199-210.
Eberlein, Robert L., David W. Peterson, and William T. Wood
1990 Casual Tracing: One Technical Solution to the Modeling Dilemma Ventana Systems, Inc.
Edmonson, Amy
1999 Psychological Safety and Learning Behavior in Work Teams. Administrative Science Quarterly 44(2):350-383.
Edwards, Paul
2000 Global Climate Science, Uncertainty, and Politics: Data-laden Models, Model-filtered Data. Science as Culture 8(4):437-472.
Ehrich, Roger W.
1983 Research on Human-Computer Interaction at Virginia Tech. SIGCHI Bulletin 14(3):5.

Enthoven, Alain C., and K. Wayne Smith
1971 How Much is Enough? Shaping the Defense Program, 1969-1969. New York: Harper and Row.
Epstein, Joshua
2006 Generative Social Science: Studies in Agent-Based Computational Modeling. Princeton, NJ: Princeton University Press.
Epstein, Joshua M.
1999 Agent-Based Computational Models and Generative Social Science. John Wiley & Sons, Inc. 4(5).
Essinger, James
2004 Jacquard's Web: How a Hand Loom Led to the Birth of the Information Age. New York: Oxford.
Fagiolo, Giorgio, Christopher Birchenhall, and Paul Windrum
2007 Empirical Validation in Agent-Based Models. Computational Economics 30(3).
Fagiolo, Giorgio, Paul Windrum, and Alessio Moneta
2006 Empirical Validation of Agent-Based Models: A Critical Survey: Laboratory of Economics and Management Sant'Anna School of Advanced Studies.
Feinstein, Andrew H., and Hugh M. Cannon
2001 Fidelity, Verifiability and Validity of Simulation: Constructs for Evaluation Developments in Business Simulation and Experiential Learning 28:57-67.
Feinstein, Andrew H., and Hugh M. Cannon
2002 Constructs of Simulation Evaluation. Simulation & Gaming 33(4):425-440.
Ferson, Scott
2004 Dependence in probabilistic modeling, Dempster-Shafer theory, and probability bounds analysis. Pp. 151. Albuquerque, NM: Sandia National Laboratories.
Ferson, Scott
2007 Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty: Sandia National Laboratories.
Fisher, Michael
1995 Representing and Executing Agent-Based Systems Workshop on Agent Theories, Architectures, and Languages. Workshop on Intelligent Agents. Amsterdam, The Netherlands, 1995, pp. 307-323. Springer-Verlag New York, Inc.
Fitzpatrick, Anne, and Ivan Olrich
2007 The Stockpile Stewardship Program: Fifteen Years On. Washington, DC: Federation of American Scientists.

Ford, Andrew, and Hilary Flynn
2005 Statistical Screening of System Dynamics Models. System Dynamics Review 21(4):273-303.
Forrester, Jay W.
1971 Counterintuitive Behavior of Social Systems Theory and Decision 2:109-140.
Forrester, Jay W., and Peter M. Senge
1980 Tests for Building Confidence in System Dynamics Models TIMS Studies in the Management Sciences 14:209-228.
Fossett, Christine A., et al.
1991 An Assessment Procedure for Simulation Models: A Case Study. Operations Research 39(5):710-723.
Frank, Ulrich, and Klaus G. Troitzsch
2005 Epistemological Perspectives on Simulation. *In* Journal of Artificial Societies and Social Simulation, Vol. 4.
Franklin, Allan
1998 Selectivity and the Production of Experimental Results "Any fool can take data. It's taking good data that counts" Arch. Hist. Exact Sci. 53:399-485.
Frokjaer, Erik, and Kasper Horkbaek
2008 Metaphors of Human thinking for Usability Inspection and Design. ACM Transactions on Computer-Human Interactions 14(4):ARM Transactions on Computer-Human Interaction.
Futral, W. Scott, et al.
1999 Performance of ALE3D on the ASCI Machines. *In* Nuclear Explosives Code Development Conference. Pp. 6. Las Vegas, Nevada: Lawrence Livermore National Laboratory.
Gass, Saul I.
1983 Decision-Aiding Models: Validation, Assessment and Related Issues for Policy Analysis. Operations Research 31(4):603-631.
Gilbert, Nigel
2004 Open Problems in Using Agent-Based Models in Industrial and Labor Dynamics Advances in Complex Systems 7(2):285-288.
Gilbert, Nigel, and Pietro Terna
1999 How to Build and use Agent-Based Models in Social Science
Gilli, M., and P. Winker
2002 A Global Optimization Heuristic for Estimating Agent Based Models
Goldstein, Harry
2006 Modeling Terrorists: New Simulators Could Help Intelligence Analysts Think Like the Enemy. IEEE Spectrum (September 2006):26-35.
Gong, Mingrui, and D. J. Murray-Smith
1998 A practical exercise in simulation model validation. Mathematical and Computer Modelling of Dynamical Systems 4(1):100-117.
González, Roberto
2007 Towards mercenary anthropology? Anthropology Today 23(3):14-19.
González, Roberto, and David Price
2007 When Anthropologists Become Counter-Insurgents. *In* Counterpunch.
Goodenough, John B., and Susan L. Gerhard
1975 Toward a Theory of Test Data Selection. IEEE Transactions on Software Engineering SE-1(2):156-173.

Greif, Siegfried

1991 Organizational Issues and Task Analysis. *In* Human Factors for Informatics Usability. B. Shackel and S. Richardson, eds. Cambridge, UK: Cambridge University Press.

Grimm, Volker

2005 Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology Science 310(5750):987-991.

Gross, Dominique, and Roger Strand

2000 Can Agent-Based Models Assist Decisions on Large-Scale Practical Problems? A Philosophical Analysis Complexity 5(6):26-33.

Grosz, Barbara J., and Candace Sidner

1986 Attention, Intentions, and the Structure of Discourse. Computational Linguistics 12(3):175-204.

Gruhl, J., and N. Gruhl

1978 Methods and Examples of Model Validation - An Annotated Bibliography. *In* MIT Energy Laboratory Working Paper.

Halverson, Christine

2001 Activity Theory and Distributed Cognition; or, What does CSCW Need to DO with Theories? Journal of Computer Supported Collaborative Work 11(1-2):24.

Hanneman, Robert A., Randall Collins, and Gabriele Mordt

1995 Discovering Theory Dynamics by Computer Simulation: Experiments on State Legitimacy and Imperialist Capitalism. Sociological Methodology 25:1-46.

Harms, William F.

1998 The Use of Information Theory in Epistemology. Philosophy of Science 65(3):472-501.

Harrison, J. Richard, and Glenn R. Carroll

1991 Keeping the Faith: A Model of Cultural Transmission in Formal Organizations. Administrative Science Quarterly 36(4):552-582.

Hayden, Nancy Kay

2007 Verification and Validation of Social Science Simulations: Some Thoughts for Consideration in National Security Applications: Defense Threat Reduction Agency/Advanced Systems and Concepts Office (DTRA/ASCO)

Hedstrom, Peter

2006 Explaining the Growth Patterns of Social Movements

Heitmeyer, Constance, James Kirby, and Bruce Labaw

1997 Tools for Formal Specification, Verification, and Validation of Requirements. COMPASS.

Helmhout, Jan Martin

2006 The Social Cognitive Actor: A multi-actor simulation of organisations

Helton, J. C.
1993 Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliability engineering & systems safety 42:327-367.
Helton, Jon Craig, and Freddie J. Davis
2000 Sampling-Based Methods for Uncertainty and Sensitivity Analysis.
Helton, J. C., and F. J. Davis
2002 Illustration of Sampling-Based Methods for Uncertainty and Sensitivity Analysis. Risk Analysis 22(3).
Helton, J. C., J. D. Johnson, and W. L. Oberkampf
2004 An Exploration of Alternative Approaches to the Representation of Uncertainty in Model Predictions Reliability Engineering and System Safety 85:39-71.
Helton, J. C., J. D. Johnson, and W. L. Oberkampf
2006 Probability of Loss of Assured Safety in Temperature Dependent Systems with Multiple Weak and Strong Links Reliability Engineering and System Safety 91:320-348.
Hemez, Francois M.
2004 The Myth of Science-Based Predictive Modeling. *In* Foundations '04 Workshop for Verification, Validation and Accreditation in the 21st Century. Pp. 41. Santa Fe, NM: Los Alamos National Laboratory.
Hendry, David F., and Neil R. Ericcson
2001 Understanding Economic Forecasts. Cambridge, MA: MIT Press.
Hermann, Charles F., and Margaret G. Hermann
1967 An Attempt to Simulate the Outbreak of World War I. The American Political Science Review 61(2):400-416.
Hills, Richard G., and Timothy Trucano
1999 Statistical Validation of Engineering and Scientific Models: Background. Albuquerque, NM: Sandia National Laboratories.
Hodges, Ann, et al.
2001 ASCI Software Quality Engineering: Goals, Principles and Guidelines. Pp. 12 with 2 appendices. Albuquerque, NM: Sandia National Laboratories.
Hodges, James S.
1991 Six (Or So) Things You Can Do with a Bad Model. Operations Research 39(3):355-365.
Hollocks, Brian
1983 Simulation and the Micro. The Journal of the Operational Research Society 34(4):331-343.
Hoppe, Thomas, and Pedro Meseguer
1993 VVT Terminology: A Proposal IEEE Expert 8(3):48-55.
Howarth, Jonathan, Terence Andre, and Rex Hartson
2007 A Structured Process for Transforming Usability Data into Usability Information. Journal of Usability Studies 3(1):7-23.
Hsiung, Pao-Ann, and Shang-Wei Lin
2006 Automatic synthesis and verification of real-time embedded software for mobile and ubiquitous systems. Embedded Systems 34(4):153-169.
Hsiung, Pao-Ann, and Shang-Wei Lin
2007 Uncertainty, probability and information-gaps. Alternative Representations of Epistemic Uncertainty 85(1-3):249-266.

Hsu, D. A., and J. S. Hunter
1977 Analysis of Simulation-Generated Responses Using Autoregressive Models. Management Science 24(2):181-190.
Humphrey, Curtis, and Julie Adams
2008 Compass Visualizations for Human-Robotic Interaction. *In* HRI'08. Amsterdam, Netherlands.
Ignall, Edward J., Peter Kolesar, and Warren E. Walker
1978 Using Simulation to Develop and Validate Analytic Models: Some Case Studies. Operations Research 26(2):237-253.
Iman, R. L., and J. C. Helton
1988 Investigation of uncertainty and sensitivity analysis techniques for computer models. Risk Anal.; Vol/Issue: 8:1:Pp: 71-90.
Irobi, Ijeoma Sandra, Johan Andersson, and Anders Wall
2004 Correctness Criteria for Models' Validation - A Philosophical Perspective. *In* International Conference on Modeling, Simulation and Visualization Methods (MSV) Las Vegas, NV.
Jacklin, Stephen A., et al.
2004 Verification, Validation, and Certification Challenges for Adaptive Flight-Critical Control System Software. *In* AIAA Guidance, Navigation, and Control Conference and Exhibit Providence, RI: AIAA.
Janssen, Marco
2005 Empirically-based agent-based models, Indiana University (Bloomington, IN), 2005.
Janssen, Marco, and Elinor Ostrom
2006 Empirically based, agent-based models. Ecology and Society 11(2):37.
Jasanoff, Sheila
1986 Risk Management and Political Culture. New York: Russell Sage Foundation.
Jasanoff, Sheila
2003 (No?) Accounting for Expertise. Science and Public Policy 30(3):157-162.
Jasanoff, Sheila
2007 Bandwagon Effects and Error Bars in particle Physics Nuclear Instruments and Methods in Physics Research A(571):704-708.
Jeng, Monwhea
2005 A selected history of expectation bias in physics. arXiv:Physics/5058199
Jennings, Nicholas R.
2000 On Agent-Based Software Engineering. Artificial Intelligence 117:277-296.
Jin, Yan, and Raymond E Levitt
1996 The Virtual Design Team: A Computational Model of Project Organizations. Computational and Mathematical Organization Theory 2(3):171-196.

Johnson, Brian, and Ben Shneiderman
1991 Tree-Maps: A Space-Filing Approach to the Visualization of Hierarchical Information Structures. IEEE:284-291.
Judd, Kenneth L.
2006 Computationally intensive analyses in economics Handbook of Computational Economics 2: Agent-Based Computational Economics.
Kamat, Vineet R., and Julio C. Martinez
2003 Validating Complex Construction Simulation Models Using 3D Visualization. Systems Analysis Modelling Simulation 43(4):455-467.
Kampmann, Christian, and John D. Sterman
1998 Feedback Complexity, Bounded Rationality, and Market Dynamics
Kaner, Cem, Jack L. Falk, and Hung Quoc Nguyen
1999 Testing Computer Software. New York John Wiley & Sons.
Kang, Hyunmo, et al.
2008 Interactive Entity Resolution in Relational Data: A Visual Analytic Tool and Its Evaluation. IEEE Transactions on Visualization and Computer Graphics (Forthcoming.).
Kelle, Udo
2005 "Emergence" vs. "Forcing" of Empirical Data? A Crucial Problem of "Grounded Theory" Reconsidered. Forum: Qualitative Social Research 6(2):Article 27.
Kelle, Udo
2005 "Emergence" vs. "Forcing" of Empirical Data? A Crucial Problem of "Grounded Theory" Reconsidered 6(2).
Kelty, Christopher
2003 Qualitative Research in the Age of the Algorithm: New Challenges in Cultural Anthropology. *In* 2003 Annual Meeting: Rethinking the Humanities in a Global Age. Boston Public Library, Boston, MA: Online Computer Library Center/Research Libraries Group, Inc (OCLC/RLG).
Kennedy, Marc C., and Anthony O'Hagan
2001 Bayesian Calibration of Computer Models Royal Statistical Society 63(3):425-464.
Kennedy, Marc C., Anthony O'Hagan, and Neil Higgins
2001 Bayesian Analysis of Computer Code Outputs.
Kerbyson, Darren J., Adolfy Hoisie, and Harvey J. Wasserman
2005 A Performance Comparison between the Earth Simulator and Other Terascale Systems on a Characteristic ASCI Workload. Concurrency and Computation: Practice and Experience 17:1219-1238.
Kieras, David E., Scott D. Wood, and David E. Meyer
1997 Predictive Engineering models Base don the EPIC Architecture for a Multimodal High-Performance Human Computer-Interaction Task. ACM Transactions on Computer-Human Interaction 4(3):230-275.
Kiewe, Howard
2008 How May I Help You? An Ethnographic View of Contact-Center HCI. Journal of Usability Studies 3(2):74-89.

Kleijnen, J. P. C.
1995 Verification and Validation of Simulation Models European Journal of Operational Research 82:145-162.
Kleijnen, J. P. C.
1999 Validation of models: statistical techniques and data availability. Winter Simulation Conference, 1999. Vol. 1, pp. 647-654.
Kleijnen, J. P. C.
2007 DASE: Design and Analysis of Simulation Experiments 1-192.
Kleijnen, J. P. C., Bert Bettonvil, and Willem Van Groenendaal
1998 Validation of Trace-Driven Simulation Models: A Novel Regression Test. Management Science 44(6):812-819.
Kleijnen, J. P. C., and J. C. Helton
1999 Statistical analyses of scatterplots to identify important factors in largescale simulations, 1: Review and comparison of techniques. Reliability engineering & systems safety 65.
Kleijnen, J. P. C., and J. C. Helton
1999 Statistical analyses of scatterplots to identify important factors in large-scale simulations, 2: robustness of techniques. Reliability engineering & systems safety 65.
Kleijnen, Jack P. C., Thomas H. Naylor, and Terry G. Seaks
1972 The Use of Multiple Ranking Procedures to Analyze Simulations of Management Systems: A Tutorial. Management Science 18(6):B245-B257.
Kleijnen, J. P. C. [1], and R. G. Sargent
2000 A methodology for fitting and validating meta-models in simulation. European Journal of Operational Research 120:14-29.
Klein, Esther E., and Paul J. Herskovitz
2005 Philosophical foundations of computer simulation validation. Simul. Gaming 36(3):303-329.
Klein, Richard, et al.
2006 ASC Predictive Science Academic Alliance Program Verification and Validation White Paper. Livermore, CA: Lawrence Livermore National Laboratory.
Kleindorfer, George B., and Ram Ganeshan
1993 The philosophy of science and validation in simulation. Los Angeles, California, United States: ACM.
Kleindorfer, George B., Liam O'Neill, and Ram Ganeshan
1998 Validation in Simulation: Various Positions in the Philosophy of Science. Management Science 44(8):1087-1099.
Klir, George
2003 An Update on Generalized Information Theory. *In* International Symposium on Imprecise Probabilities and Their Applications. J. M. Bernard, T. Seidenfeld, and M. Zaffalon, eds. Lugano, Switzerland: Carleton Scientific.
Knepell, Peter L., and Deborah Arangno
1993 Simulation Validation - a Confidence Assessment Methodology: IEEE Computer Society Press.

Knupp, Patrick M., and Kambiz Salari
2003 Verification of Computer Codes in Computational Science and Engineering: CRC Press.
Konikow, Leonard F., and John D. Bredehoeft
1992 Ground-water Models Cannot be Validated. Advances in Water Resources 15:75-83.
Kotler, Philip, and Randall L. Schultz
1970 Marketing Simulations: Review and Prospects. The Journal of Business 43(3):237-295.
Krahmer, Elizabeth M.
1997 Everything you Ever Wanted to Know about critiquing Models, But Were Afraid to Ask
Kramer, Adam D. I., and Kerry Rodden
2007 Applying a User-Centered Metric to Identify Active Blogs. *In* CHI 2007. Pp. 2525-2530, Vol. ACM. San Jose, CA.
Krovi, Ravindra, Arthur C. Graesser, and William E. Pracht
1999 Agent Behaviors in Virtual Negotiation Environments. IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews 29(1):15-25.
Kruger, Justin, and David Dunning
1999 Unskilled and Unaware of It: how Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self Assessments. Journal of Personality and Social Psychology 77(6):1121-1134.
Kuppers, Gunter, and Lenhard Johannes
2005 Validation of Simulation: Patterns in the Social and Natural Sciences Journal of Artificial Societies and Social Simulation 8(4).
Kuznar, Lawrence
1997 Reclaiming a Scientific Anthropology. Walnut Creek, CA: Alta Mira.
Kuznar, Lawrence A.
2006 High-Fidelity Computational Social Science in Anthropology: Prospects for Developing a Comparative Framework. Social Science Computer Review 24(1):1-15.
Lahsen, Myanna
2005 Seductive Simulations? Uncertainty Distribution Around Climate Models. Social Studies of Science 35(6):895-922.
Lakoff, George, and Mark Johnson
1980 Metaphors We Live By. Chicago, IL: University of Chicago.
Langley, Paul A., Mark Paich, and John D. Sterman
1998 Explaining Capacity Overshoot and Price War: Misperceptions of Feedback in Competitive Growth Markets. *In* 1998 International System Dynamics Conference.
Lansing, J. Stephen
2002 Artificial Societies: and the Social Sciences. Artificial Life 8:279-292.
Lansing, J. Stephen
2003 Complex Adaptive Systems. Annual Review of Anthropology 32:183-294.

Lansing, J. Stephen, and James N. Kremer

1993 Emergent Properties of Balinese Water Temple Networks: Co-adaptation on a Rugged Fitness Landscape. American Anthropologist 95:97-114.

Laughlin, Robert

2002 The Physical Limits of Computability. Computing in Science and Engineering (May/June):27-30.

Lauw, Hady, et al.

2005 Social Network Discovery by Mining Spatio-Temporal Events. Computational and Mathematical Organization Theory 11:97-118.

Law, Averill M.

2006 How to Build Valid and Credible Simulation Models. Winter Simulation Conference:58-66.

Law, Averill M., and W. David Kelton

1997 Simulation Modeling and Analysis: McGraw-Hill Higher Education.

Law, Averill M., and Michael G. McComas

2001 Building valid models: how to build valid and credible simulation models. Arlington, Virginia: IEEE Computer Society.

Lawrence, Michael, et al.

2006 Judgmental Forecasting: A Review of Progress over the last 25 years. International Journal of Forecasting 22:493-518.

Lee, Ju-Sung, Kathleen Carley, and Judith Effken

2003 Validating a Computational Model of Decision-Making using Empirical Data. NAACSOS Conference 2003, Pittsburgh, PA, 2003. Vol. Day 1.

Lempert, Robert J.

2002 A New Decision Sciences for Complex Systems. PNAS 99(3):7309-7313.

Leombruni, Roberto, et al.

2006 A Common Protocol for Agent-Based Social Simulation. Journal of Artificial Societies and Social Simulation 9(1).

Lewis, Robert O.

1992 Independent verification & validation: a life cycle engineering process for quality software: John Wiley & Sons, Inc.

Light, Jennifer

1999 When Computers Were Women. Technology and Culture 40(3):455-483.

Lin, Zhiang, and Kathleen Carley

2002 Organizational Design and Adaptation in Response to Crises: Theory and Practice.

Liu, Fei, and Ming Yang

2004 Verification and Validation of AI Simulation Systems. *In* Proceedings of the Third International Conference on Machine Learning and Cybernetics. Shanghai, China: IEEE.

Liu, Fei, Ming Yang, and Sun Guobing

2007 Verification of Human Decision Models in Military Simulations. Modelling & Simulation, 2007. AMS '07. First Asia International Conference on 2007, pp. 363-368.

Liu, Fei, Ming Yang, and Zicai Wang
2005 Study on Simulation Credibility Metrics. *In* Winter Simulation Conference. M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, eds. Pp. 2554-2560. Orlando, FL: IEEE Conference Proceedings.

Logan, Roger, and Cynthia Nitta
2002 Verification and Validation Methodology and Quantitative Reliability at Confidence: Basis for an Investment Strategy. Pp. 109. Livermore, CA: Lawrence Livermore National Laboratory.

Louie, Marcus, et al.
2003 Model Comparisons: Docking OrgAhead and SimVision. NAACSOS Conference 2003, Pittsburgh, PA, 2003. Vol. Day 3.

Mackenzie, Donald
1995 The Automation of Proof: A Historical and Sociological Exploration. IEEE Annals of the History of Computing 17(3):7-29.

Mackenzie, Donald.
2001 Mechanizing Proof: Computing, Risk and Trust. Cambridge, MA: MIT Press.

Mackerrow, Edward P.
2003 Understanding Why: Dissecting Radical Islamic Terrorism with Agent-Based Simulation. Los Alamos Science:184-191.

Macy, Michael W.
1991 Learning to Cooperate: Stochastic and Tacit Collusion in Social Exchange. The American Journal of Sociology 97(3):808-843.

Macy, Michael W., and Robert Willer
2002 From Factors to Actors: Computational Sociology and Agent-Based Modeling. Annual Review of Sociology 28:143-166.

Mahadevan, Sankaran, and Ramesh Rebba
2004 Validation of reliability computational models using Bayesian networks. 87(2):223-232.

Malerba, Franco, et al.
1999 “History-Friendly” Models of Industry Evolution: The Computer Industry Industrial and Corporate Change 8(1):3-41.

Manville, B.
1999 Complex adaptive knowledge management: a case from
McKinsey & Company. *In* The Biology of Business: Decoding the Natural Laws of Enterprise. J. H. C. III, ed. Pp. 89-111. San Francisco: Jossey-Bass.

Marchionini, Gary, and John Sibert
1991 An Agenda for Human-Computer Interaction: Science and Engineering Serving Human Needs. SIGCHI Bulletin 23(4):17-32.

Marcus, George, and Michael M. J. Fischer
1999 Anthropology as Cultural Critique: An Experimental Moment in the Human Sciences. Chicago, IL: University of Chicago.

Martin, Dale E., et al.
2004 Scheduling Optimization on the Simbus Backplane. *In* Proceedings of the 37th Annual Simulation Symposium (ANSS'04). Pp. 7: IEEE Computer Society.

Martis, Morvin Savio
2006 Validation of Simulation Based Models: A Theoretical Outlook. The Electronic Journal of Business Research Methods 4(1):39-46.

Mass, Nathaniel J.
1991 Diagnosing Surprise Model Behavior: a Took for Evolving Behavioral and Policy Insights. System Dynamics Review 7(1):68-86.

Masuch, Michael, and Perry LaPotin
1989 Beyond Garbage Cans: An Al Model of Organizational Choice. Administrative Science Quarterly 34(1):38-67.

McFate, Montgomery
2005 The Military Utility of Understanding Adversary Culture. Joint Forces Quarterly (38):42-48.

Menand, Louis
2002 What Comes Naturally: Does Evolution Explain Who We Are? *In* The New Yorker.

Menzies, Tim
2002 Verification and Validation in Artificial Intelligence. Foundations 02: A V&V Workshop, Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland, USA, 2002, pp. 70.

Menzies, Tim, and Charles Pecheur
2005 Verification and Validation and Artificial Intelligence. Advances in Computing 65:154-203.

Metz, Michael L., and Jack Jordan
2001 Verification of object-oriented simulation designs. Winter Simulation Conference, Arlington, Virginia, 2001, pp. 600-603. IEEE Computer Society.

Meyer, Mary, and Jane Booker
1991 Eliciting and Analyzing Expert Judgment: A Practical Guide. Philadelphia: SIAM.

Mihram, Arthur G.
1972 Some Practical Aspects of the Verification and Validation of Simulation Models. Operational Research Quarterly 23(1):17-29.

Millar, Alan
1984 Veridicality: More on Searle. Analysis 45(2):120-124.

Mody, Cyrus
2001 A Little Dirt Never Hurt Anyone: Knowledge Making and Contamination in Materials Science. Social Studies of Science 31(1):7-36.

Moradi, Farshad, and Rassul Ayani
2003 Parallel and distributed simulation. *In* Applied system simulation: methodologies and applications. Pp. 457-486: Kluwer Academic Publishers.

Morecroft, John D. W.
1985 Rationality in the Analysis of Behavioral Simulation Models. Management Science 31(7):900-916.

Morse, Emile, and Michelle Potts Steves
2000 CollabLogger: A Tool for Visualizing Groups at Work. Gaithersburg, MD: National Institute of Standards and Technology.

Moss, Scott
2008 Alternative Approaches to the Empirical Validation of Agent-Based Models JASSS 11.
Moss, Scott, Bruce Edmonds, and Steve Wallis
1997 Validation and Verification of Computational Models with Multiple Cognitive Agents: Manchester Metropolitan University
Mulligan, Kevin, and Barry Smith
1986 A Husserlian Theory of Indexicality. Grazer Philosophische Studien 28:133-63.
Na Alessa, Lilian, Melinda Laituri, and Michael Barton
2006 An "All Hands" Call to the Social Science Community: Establishing a Community Framework for Complexity Modeling Using Agent Based Models and Cyberinfrastructure. *In* Journal of Artificial Societies and Social Simulation, Vol. 9.
Nance, Richard E., and Robert G. Sargent
2002 Perspectives on the Evolution of Simulation. Operations Research 50(1):161-172.
Narayanan, Srini, and Sheila A. McIlraith
2002 Simulation, Verification and Automated Composition of Web Services. WWW, Honolulu, Hawaii, 2002.
Nardi, Bonnie
1995 Studying Context: A Comparison of Activity Theory, Situated Action Models, and Distributed Cognition. *In* Context and consciousness: activity theory and human-computer interaction. B.A. Nardi, ed. Pp. 70-102. Cambridge, MA: MIT Press.
NASA
1999. Independent Verification and Validation Overview. Fairmont, WV: NASA.
Nayani, Nirupama, and Mansooreh Mollaghesemi
1998 Validation and Verification of the Simulation Model of a Photolithography Process in Semiconductor Manufacturing. Winter Simulation Conference, 1998, pp. 1071-1022.
Naylor, Thomas H.
1970 Policy Simulation Experiments with Macroeconometric Models: The State of the Art. American Journal of Agricultural Economics 52(2):263-271.
Naylor, Thomas H., Donald S. Burdick, and W. Earl Sasser
1967 Computer Simulation Experiments with Economic Systems: The Problem of Experimental Design. Journal of the American Statistical Association 62(320):1315-1227.
Naylor, Thomas H., et al.
1967 Verification of Computer Simulation Models. Management Science 14(2):B92-B106.
Naylor, Thomas H., Terry G. Seaks, and D. W. Wichern
1972 Box-Jenkins Methods: An Alternative to Econometric Models. International Statistical Review / Revue Internationale de Statistique 40(2):123-137.

Naylor, Thomas H., Kenneth Wertz, and Thomas H. Wonnacott
1969 Spectral Analysis of Data Generated by Simulation Experiments with Econometric Models. Econometrica 37(2):333-352.
Naylor, Thomas H., and Thomas H. Wonnacott
1970 A Comment on the Analysis of Data Generated by Simulation Experiments. Management Science 17(3):233-235.
Newsome, Mary R., and P. N. Johnson-Laird
2006 How Falsity Dispels Fallacies Psychology Press 12(2):214-234.
Nichols, A. L., and R. J. Zeckhauser
1988 The perils of prudence: how conservative risk assessments distort regulation. Regul-Toxicol-Pharmacol 8(1):61-75.
Nijim, Mais, Tao Xie, and Xiao Qin
2005 Performance Analysis of an Admission Controller for CPU- and I/O-Intensive Applications in Self-Managing Computer Systems. ACM SIGOPS Operating Systems Review 39(4 (October 2005)):37-45.
Nitta, Cynthia, and Roger Logan
2004 Qualitative and Quantitative Linkages from V&V to Adequacy, Certification, and Benefit/Cost Ratio. Pp. 87. Livermore, CA: Lawrence Livermore National Laboratory.
Nowak, David A., and Ronald A Christensen
1997 ASCI Applications. *In* Supercomputing 1997. Pp. 12. San Jose, CA: Lawrence Livermore National Laboratory.
O'Hagan, A.
2006 Bayesian Analysis of Computer Code Outputs: A Tutorial Reliability Engineering and System Safety 91:1290-1300.
O'Keefe, Robert M., and Daniel O'Leary
1993 Expert System Verification and Validation: A Survey and Tutorial Artificial Intelligence Review 7:3-42.
O'Leary, Daniel
1997 Validation of Computational Models Based on Multiple Heterogeneous Knowledge Sources. Computational and Mathematical Organization Theory 3(2):75-90.
Oakley, Jeremy E., and Anthony O'Hagan
2004 Probabilistic Sensitivity Analysis of Complex models: A Bayesian Approach. Journal of the Royal Statistical Society 66(3):751-769.
Oberkampf, William, and Timothy Trucano
2004 Design of and Comparison with Verification and Validation Benchmarks. Albuquerque, NM: Sandia National Laboratories.
Oberkampf, William, Timothy Trucano, and Charles Hirsch
2003 Verification, Validation and Predictive Capability in Computational Engineering and Physics. Albuquerque, NM: Sandia National Laboratories.
Oberkampf, William L., and Matthew F. Barone
2005 Measures of Agreement Between Computation and Experiment: Validation Metrics Sandia National Laboratories

Oberkampf, W. L., F. G. Blottner, and D. P. Aeschliman
1995 Methodology for Computational Fluid Dynamics Code Verification/Validation. AIAA Fluid Dynamics Conference, San Diego, CA, 1995.
Oberkampf, William L., Pilch Martin, and Timothy G. Trucano
2007 Predictive Capability Maturity Model for Computational Modeling and Simulation. Albuquerque, NM: Sandia National Laboratories.
Oberkampf, W. L., and Timothy Trucano
2002 Verification and Validation in Computational Fluid Dynamics Sandia National Laboratories
Oberkampf, William L., and Timothy Trucano
2006 Design of and Comparison with Verification and Validation Benchmarks
Oberkampf, William L., and Timothy Trucano
2007 Verification and Validation Benchmarks. Pp. 67. Albuquerque, NM: Sandia National Laboratories.
Oberkampf, William L., Timothy G. Trucano, and Charles Hirsch
2004 Verification, validation, and predictive capability in computational engineering and physics. Applied Mechanics Reviews 57(5):345-384.
Office of Defense Programs, United States Department of Energy
1995 (May) The Stockpile Stewardship and Management Program: Maintaining Confidence in the Safety and Reliability of the Enduring US Nuclear Weapon Stockpile: United States Department of Energy.
Office of Science, United States Department of Energy
4 March 2000 Scientific Discovery Through Advanced Computing. Department of Energy, ed. Washington, DC.
Oliva, Rogelio
2002 Model Calibration as a Testing Strategy for System Dynamics Models. European Journal of Operations Research 151:552-568.
Oren, Tuncer I.
1981 Concepts and criteria to assess acceptability of simulation studies: a frame of reference. Commun. ACM 24(4):180-189.
Oreskes, Naomi
1996 The Role of Quantitative Models in Science. *In* Models in Ecosystem Science. C. D. Canham, J. J. Cole, and W. K. Lauenroth, eds. Pp. 13-31. Princeton: Princeton University Press.
Oreskes, Naomi
1998 Evaluation (not validation) of Quantitative Models. Environ Heath Perspect 106(6):1453-1460.
Oreskes, Naomi
2003 Why Believe a Computer? Models, Measures, and Meaning in the Natural World. *In* The Earth Around Us: Maintaining a Livable Planet. Pp. 70-82. Boulder, CO: Westview.
Oreskes, Naomi, Kristen Shrader-Frechette, and Kenneth Belitz
1994 Verification, Validation and Confirmation of Numerical Models in the Earth Sciences. Science 263(5147):641-646.

Orlikowski, Wanda J., and Debra C. Gash
1994 Technological Frames: Making Sense of Information Technology in Organizations. ACM Transactions on Information Systems 12(2):174-204.
Overstreet, C. Michael
2002 Improving the model development process: model testing: is it only a special case of software testing? San Diego, California: Winter Simulation Conference.
Pace, Dale K.
1993 Naval Modeling and Simulation Verification, Validation and Accreditation Winter Simulation Conference 1993.
Pace, Dale K.
2004 Modeling and Simulation Verification and Validation Challenges Johns Hopkins Apl Technical Digest 25(2).
Page, Ernest H., Bradford S. Canova, and John A. Tufarolo
1997 A Case Study of Verification, Validation and Accreditation for Advanced Distributed Simulation. ACM Transactions on Modeling and Computer Simulation (TOMACS):393-424.
Panzarasa, P., and Kathleen Carley
2003 Multi-Agent Negotiation: Logical Foundations and Computational Complexity. NAACSOS Conference 2003, 2003. Vol. Day 3.
Park, Woojin, Don B. Chaffin, and Bernard J. Martin
2004 Toward Memory-Based Human Motion Simulation: Development and Validation of a Motion Modification Algorithm. IEEE Transactions on Systems, Man and Cypernetics - Part A: Systems and Humans 34(3):376-385.
Parry, G. W., and P. W. Winter
1981 Characterization and evaluation of uncertainty in probabilistic risk analysis. Nucl. Saf.; Vol/Issue: 22:1:Pp.: 28-42.
Pearl, Judea.
1987 Evidential reasoning using stochastic simulation of causal models. Artificial Intelligence 32:245-257.
Pecheur, Charles
2006 Verification of Intelligent Control Software. Proceedings of the 9th ESA Workshop on Advanced Space Technologies for Robotics and Automation (ASTRA 2006), Noordwijk, The Netherlands, 2006.
Penner, David E.
2000-2001 Cognition, Computers, and Synthetic Science: Building Knowledge and Meaning through Modeling. Review of Research in Education 25:1-35.
Peterson, David W., and Robert L. Eberlein
1994 Reality Check: a Bridge Between Systems Thinking and Systems Dynamics System Dynamics Review 10(2-3):159-174.
Pfaffenberger, Bryan
1992 Social Anthropology of Technology. Annual Review of Anthropology 21:491-516.

Picard, Richard R., and R. Dennins Cook
1984 Cross-Validation of Regression Models. Journal of the American Statistical Association 79(387):575-583.
Pilch, Martin, et al.
2004 The Role of V&V in QMU. *In* Foundations -04 Workshop for Verification, Validation and Assessment in the 21st Century. Pp. 37. Tempe, AZ.
Pilch, Martin, et al.
2001 Guidelines for Sandia ASCI Verification and Validation Plans - Content and Format: Version 2.0. Pp. 82. Albuquerque, NM: Sandia National Laboratories.
Pilch, Martin L, et al.
2004 Concepts for Stockpile Computing (OUO Report). Pp. 109. Albuquerque, NM: Sandia National Laboratories.
Pilkey, Orrin, and Linda Pilkey-Jarvis
2007 Useless Arithmetic: Why Environmental Scientists Can't Predict the Future Nature 447.
Pinch, Trevor
1993 "Testing - One, Two, Three... Testing!": Toward a Sociology of Testing Science, Technology and Human Values 18(1):25-41.
Plaisant, Catherine, Jean Daniel Fekete, and Georges Grinstein
2008 Promoting Insight-Based Evaluation of Visualizations: From Contest to Benchmark Repository. IEEE Transactions on Visualization and Computer Graphics 14(1):120-134.
Plaisant, Catherine, and Ben Shneiderman
2005 Show Me! Guidelines for Producing Recorded Demonstrations. *In* IEEE Symposium on Visual Languages and Human Centric Computing (VL-HCC'05. Pp. 8.
Polhill, Gary J., Luis R. Izquierdo, and Nicholas M. Gotts
2005 The ghost in the model (and other effects of floating point arithmetic). Journal of Artificial Societies and Social Simulation 8(1).
Poolla, Kameshwar, et al.
1994 A Time-Domain Approach to Model Validation IEEE Transactions on Automatic Control 39(5):951-959.
Post, Douglass E., and Richard P. Kendall
2004 Software Project Management and Quality Engineering: Practices for Complex, Coupled Multiphysics, Massively Parallel Computational Simulations: Lessons Learned from ASCI. The International Journal of High Performance Computing Applications 18(4):17.
Post, Douglass E., and Lawrence G. Votta
2005 Computational Science Demands a New Paradigm. American Institute of Physics.
Putnam, Hilary
1980 Models and Reality. The Journal of Symbolic Logic 45(3):464-482.
Pygott, C. H., et al.
1992 Will Proof Replace Simulation? [and Discussion]. Philosophical Transactions: Physical Sciences and Engineering 339(1652):21-33.
Qu, Lijun, and Gary W. Meyer
2006 Perceptually Driven Interactive Geometry Remeshing. *In* I3D 2006. Redwood City, CA: ACM.

Qudrat-Ullah, Hassan
2005 Structural Validation of System Dynamics and Agent-Based Simulation Models Proceedings 19th European Conference on Modelling and Simulation, 2005.
Rahmandad, Hazhir, and John D. Sterman
2006 Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-Based and Differential Equation Models: MIT Sloan School of Management.
Rastetter, Edward B.
1996 Validating Models of Ecosystem Response to Global Change. BioScience 46(3):190-198.
Ray, Daryll, and Earl O. Heady
1972 Government Farm Programs and Commodity Interaction: A Simulation Analysis. American Journal of Agricultural Economics 54(4):578-590.
Raybourn, Elaine M., and Chris Forsythe
2001 Toward the Computational Representation of Individual Cultural, Cognitive, and Physiological State: The Sensor Shooter Simulation Sandia National Laboratories
Rebba, Ramesh, Sankaran Mahadevan, and Shuping Huang
2005 Validation and Error Estimation of Computational Models Los Alamos National Laboratory.
Redmond, Jim
2003 VV-5.1 FY2005 Q1: Hostile Mechanical. Pp. 5: Sandia
Reid, Richard A., and Elisa L. Koljonen
1999 Validating a Manufacturing Paradigm: A System Dynamics Modeling Approach Winter Simulation Conference 1999.
Reminga, Jeffrey, and Kathleen Carley
2003 Measures for ORA (Organizational Risk Analyzer).
Ren, Yuqing, Kathleen Carley, and L. Argote
2001 Simulating the Role of Transitive Memory in Group Training and Performance. CASOS Conference 2001, Pittsburgh, PA, 2001. Vol. Day 3.
Risbey, James, Milind Kandlikar, and Anand Patwardhan
1996 Assessing Integrated Assessments. Climatic Change 34:369-395.
Roache, Patrick J.
1998 Verification and Validation in Computational Science and Engineering. Albuquerque, NM: Hermosa Publishing,
Robinson, Stewart
1997 Simulation Model Verification and Validation: Increasing the Users' Confidence Winter Simulation Conference 1997.
Robinson, Stewart
1999 Three Sources of Simulation Inaccuracy (and how to overcome them). Winter Simulation Conference, 1999, pp. 1707-1708.

Rosenhead, J. V.
1968 Experimental Simulation of a Social System. OR 19(3):289-298.
Rosenstock, Linda, and Lore Jackson Lee
2002 Attacks on Science: The Risks to Evidence-Based Policy. Ethics and Public Health 92(1):14-18.
Roth, Camille
2007 Empiricism for Descriptive Social Network Models. Physica A 378:53-58.
Rouchier, Juliette
2005 Data gathering to build and validate small-scale social models for simulation. Two ways: strict control and stake-holders involvement: GREQAM.
Rykiel, Edward J., Jr.
1996 Testing Ecological Models: the meaning of validation. Ecological Modelling 90:229-244.
Saito, Hitomi, and Kazuhisa Miwa
2002 A Cognitive Study of Information Seeking Processes in the WWW: The Effects of Searchers' Knowledge and Experience. IEEE ??? :321-327.
Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts
2006 Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. Science 311(10 February 206):854-856.
Sallans, Brian, et al.
2003 Simulation and Validation of an Integrated Markets Model JASSS 6(4).
Sandoval, D. L.
1987 CAVEAT calculations of shock interactions. United States.
Sarewitz, Daniel, Roger A. Pielke, Jr., and Radford Byerly, Jr.
2000 Prediction: Science, Decision Making and the Future of Nature. Washington, DC: Island Press.
Sargent, Robert G.
1979 Validation of simulation models. Winter Simulation Conference, San Diego, CA, United States, 1979, pp. 497-503. IEEE Press.
Sargent, Robert G.
1985 An Expository on Verification and Valification of Simulation Models. Winter Simulation Conference, 1985, pp. 15-22.
Sargent, Robert G.
1987 An Overview of Verification and Validation of Simulation Models. Winter Simulation Conference, 1987, pp. 33-39.
Sargent, Robert G.
1988 A tutorial on validation and verification of simulation models Winter Simulation Conference San Diego, California, United States, 1988, pp. 33-39 ACM.
Sargent, R. G.
1991 Simulation Model Verification and Validation Winter Simulation Conference Phoenix, AZ, 1991, pp. 37-47.
Sargent, Robert G.
1998 Verification and Validation of Simulation Models Winter Simulation Conference, 1998.

Sargent, Robert G.
2000 Verification, Validation and Accreditation of Simulation Models. Winter Simulation Conference, 2000. Vol. 1, pp. 50-59.
Sargent, Robert G.
2001 Verification and Validation: Some Approaches and Paradigms for Verifying and Validating Simulation Models. Winter Simulation Conference, Arlington, VA, 2001. Vol. 1, pp. 106-114.
Sargent, Robert G.
2005 Verification, Validation and Accreditation of Simulation Models. Winter Simulation Conference, 2005.
Sargent, Robert G., et al.
2000 Strategic directions in Verification Validation and Accreditation; research: strategic directions in Verification, Validation, and Accreditation research. Winter Simulation Conference, Orlando, Florida, 2000, pp. 909-916. Society for Computer Simulation International.
Sarkar, Palash
2000 A Brief History of Cellular Automata. ACM Computing Surveys 32(1):80-107.
Saysel, Ali Kerem, and Yaman Barlas
2004 Model Simplification and Validation: Illustration with Indirect Structure Validity Tests Working Papers in System Dynamics.
Schmid, Alex
2005 What is the Truth of Simulation? JASSS 8(4).
Schrank, William E., and Charles C. Holt
1967 Critique of: "Verification of Computer Simulation Models". Management Science 14(2):B-104-B-106.
Schreiber, Craig, and Kathleen Carley
2003 The Impact of Databases on Knowledge Transfer: Simulation Providing Theory. NAACSOS Conference 2003, Pittsburgh, PA, 2003. Vol. Day 2.
Shackel, Brian, and Brian Richardson
1991 Human Factors for Informatics Usability. Cambridge, UK: Cambridge University Press.
Shackley, Simon, and Brian Wynne
1995 Integrating Knowledge for Climate Change: Pyramids, Nets and Uncertainties. Global Environmental Change 1995:113-126.
Shackley, Simon, and Brian Wynne
1996 Representing Uncertainty in Global Climate Change Science and Policy: Boundary-Ordering Devices and Authority. Science, Technology and Human Values 21(3):275-302.
Shackley, Simon, et al.
1998 Uncertainty, Complexity and Concepts of Good Science in Climate Change Modelling: Are GCMs the Best Tools? Climatic Change 38:195-205.
Shankar, Anurag, et al.
2002 Building and Supporting a Massive Data Infrastructure for the Masses. *In* SIGUCCS 2002. Providence, Rhode Island: ACM.

Shannon, Robert E.
1981 Tests for Verification and Validation of Computer Simulation Models. Winter Simulation Conference 2.
Shannon, Robert E.
1992 Introduction to Simulation. Winter Simulation Conference, 1992, pp. 65-73.
Shanthikumar, J. G., and Sargent, R. G.
1983 A Unifying View of Hybrid Simulation/Analytic Models and Modeling. Operations Research 31(6):1030-1052.
Shapiro, Harold T.
1973 Is Verification Possible? The Evaluation of Large Econometric Models. American Journal of Agricultural Economics 55(2):250-258.
Shi, Xiaolin, Lada A. Adamic, and Martin J. Strauss
2007 Networks of Strong Ties. Physica A 378:33-47.
Shneiderman, Ben
1979 Human Factors Experiments in Designing Interactive Systems. IEEE Transactions (??).
Shneiderman, Ben
1991 Human Values and the Future of Technology: A Declaration of Responsibility. SIGCHI Bulletin 23(1):6.
Shneiderman, Ben
1996 The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization. *In* Proceedings of the IEEE Conference on Visual Languages, Vol. 336-343: IEEE.
Shneiderman, Ben
1998 Designing the User Interface: Strategies for Effective Human-Computer Interaction. Reading, MA: Addison-Wesley.
Shneiderman, Ben, et al.
2005 Turning Information Visualization Innovations into Commercial Products: Lessons to Guide the Next Success. *In* IEEE Symposium on Information Visualization. Minneapolis, MN: IEEE.
Sidner, Candace
1978 The Use of Focus as a Tool for Disambiguation of Definite Noun Phrases. TINLAP-2, ACM and ACL:86-95.
Sidner, Candace, and Myroslava Dzikivska
2002 Hosting Activities: Experience with and Future Directions for a Robot Agent Host. IUI, San Francisco, CA, 2002. ACM.
Sidner, Candace L., et al.
2004 Where to Look: A Study of Human-Robot Engagement. *In* IUI 2004. Madeira, Funchal, Portugal: ACM.
Silverman, Barry G., et al.
2007 Modeling factions for "effects based operations": part I: Leaders and followers. Computational and Mathematical Organization Theory 13:379-406.
Sims, Benjamin
2005 Safe Science: Material and Social Order in Laboratory Work. Social Studies of Science 35(3):333-366.
Sismondo, Sergio
1999 Models, Simulations and Their Objects. Science in Context 12(2):247-260.

Situngkir, Hokky
2003 Emerging the Emergence Sociology: The Philosophical Framework of Agent-Based Social Studies Journal of Social Complexity 2.
Skias, Spiro T.
2006 Background of the Verification and Validation of Neural Networks. *In* Methods and Procedures for the Verification and Validation of Artificial Neural Networks. B.J. Taylor, ed. Pp. 1-12 New York: Springer.
Smith, Roy, et al.
1997 Model Validation for Dynamically Uncertain Systems Mathematical Modelling of Systems 3(1):43-58.
Smith, Roger D.
1998 Essential Techniques for Military Modeling and Simulation. Winter Simulation Conference, 1998, pp. 8005-812.
Smith, Roy S., and John C. Doyle
1992 Model Validation: A Connection Between Robust Control and Identification IEEE Transactions on Automatic Control 37(7):942-952.
Snee, Ronald D.
1977 Validation of Regression Models: Methods and Examples. Technometrics 19(4):415-428.
Soo, Jinwook, and Ben Shneiderman
206 Knowledge Discovery in High Dimensional Data: Case Studies and a User Survey for the Rank-by Feature Framework. IEEE Transactions on Visualization and Computer Graphics 12(3):311-322.
Soudah, Jamileh, et al.
2006 ASC V&V Program Strategy: "Toward a Predictive Enterprise": National Nuclear Security Administration.
Spiegelman, Donna, Bernard Rosner, and Roger Logan
2000 Estimation and Inference for Logistic Regression with Covariate Misclassification and Measurement Error in Main Study/Validation Study Designs. Journal of the American Statistical Association 95(449):51-61.
Sprigg, James, and Mark Ehlen
2005 Dynamic Nash Equilibrium in an Overlapping-Generations Simulation. Pp. 30. Albuquerque, NM: Sandia National Laboratories.
Srbljinovic, Armano, and Ognjen Skunca
203 An Introduction to Agent Based Modelling and Simulation of Social Processes. Interdisciplinary Description of Complex Systems 1(1-2):1-8.
Sterman, John D.
1987 Testing Behavioral Simulation Models by Direct Experiment. Management Science 33(12):1572-1592.
Sterman, John D
1991 A Skeptic's Guide to Computer Models. *In* Managing a Nation: The Microcomputer Software Catalog. G.O. Barney, ed. Pp. 209-229. Boulder, CO: Westview Press.
Sterman, John D.
2002 All Models are Wrong: Reflections on Becoming a Systems Scientist System Dynamics Review 18(4):501-531.
Sterman, John D., et al.
1994 The Meaning of Models. Science 264(5157):329-331.
Stern, Fred, Hugh W. Coleman, and Eric G. Paterson

2001 Comprehensive Approach to Verification and Validation of CFD Simulations—Part 1: Methodology and Procedures. Journal of Fluids Engineering 123(4):793-802.

Stevenson, D. E.
1994 Science, Computational Science, and Computer Science: At a Crossroads. Communications of the ACM 37(12):84-96.

Steves, Michelle Potts, and Emile Morse
2001 Mining Usability Information from Log Files: A Multi-Pronged Approach. Gaithersburg, MD: National Institute of Standards and Technology.

Steves, Michelle Potts, and Emile Morse
2001? Looking at the Whole Picture: A Case Study of Analyzing a Virtual Workplace. Gaithersburg, MD: National Institute of Standards and Technology.

Stewart, Graig. A., et al.
2001 High Performance Computing: Delivering Valuable and Valued Services at Colleges and Universities. SIGUCCS, Portland, OR, 2001. Vol. User Services Conference; Vol. 29 archive.

Proceedings of the 29th annual ACM SIGUCCS conference on User services, pp. 266 - 269. ACM.

Stewart, Thomas R., and Cynthia M. Lusk
1994 Seven Components of Judgmental Forecasting Skill: Implications for Research and the Improvement of Forecasts. Journal of Forecasting 13:579-599.

Stone, M.
1974 Cross-Validation and Multinomial Prediction. Biometrika 61(3):509-515.

Stubblefield, William
2007 Toward a Situated Theory of Computer Models. Pp. 8. Albuquerque, NM: Sandia National Laboratories.

Suchman, Lucy
2000 Located Accountabilities in Technology Production: Centre for Science Studies, Lancaster University, Lancaster, UK.

Sweeney, Linda Booth, and John D. Sterman
2000 Bathtub Dynamics: Initial results of a Systems Thinking Inventory System Dynamics Review 16(4):249-286.

Swiler, Laura P.
2005 Gaussian Processes in Response Surface Modeling. *In* SAND2005-6892. Pp. 1-9: Sandia National Laboratories.

Swiler, Laura P., and Timothy Trucano
2005 Calibration Under Uncertainty Sandia National Laboratories.

Takadama, Keiki, et al.
2003 Cross-Element Validation in Multiagent-based Simulation: Switching Learning Mechanisms in Agents. *In* Journal of Artificial Societies and Social Simulation, Vol. 6.
Taylor, A. J.
1983 The Verification of Dynamic Simulation Models. The Journal of the Operational Research Society 34(3):233-242.
Taylor, Peter
1995 Building on Construction: An Exploration of Heterogeneous Constructionism, Using an Analogy from Psychology and a Sketch From Socioeconomic Modeling. Perspectives on Science 3(1):66-95.
Thacker, Ben H.
2003 The Role of Nondeterminism in Verification and Validation of Computational Solid Mechanics Models.
Theofanos, Mary, and Whitney Quesenbery
2005 Towards the Design of Effective Formative Test Reports. Journal of Usability Studies 1(1):27-45.
Thomsen, Jan, et al.
1999 A Trajectory for Validating Computational Emulation Models of Organizations. Computational and Mathematical Organization Theory 5(4):385-401.
Tooby, John, and Leda Cosmides
1992 The Psychology Foundations of Culture The Adapted Mind: Evolutionary Psychology and the Generation of Culture
Troitzsch, Klaus G.
2004 Validating Simulation Models. 18th European Simulation Multiconference, 2004.
Trucano, Timothy
1998 Prediction and Uncertainty in Computational Modeling of Complex Phenomena: A Whitepaper Sandia National Laboratories
Trucano, Timothy
2004 "Uncertainty Quantification and the Department of Homeland Security" Sandia National Laboratories.
Trucano, Timothy
2005 Uncertainty in Verification and Validation: Recent Perspectives. *In* SIAM 2005 CS&E Meeting. Pp. 45. Albuquerque, NM: Sandia National Laboratories.
Trucano, Timothy
2005 From Pencils to Computers. Albuquerque, NM: Sandia National Laboratories.
Trucano, Timothy
2006 QMU Notes for SAG presentation. Albuquerque, NM: Sandia National Laboratories.
Trucano, Timothy, et al.
2001 Description of the Sandia Validation Metrics Project: Sandia National Laboratories.

Trucano, Timothy, and Laura McNamara
2005 Notes for Cognition, Uncertainty and Decisions (CUD). Internal Sandia Document, contact authors for copy (tgtruca@sandia.gov).
Trucano, Timothy, and Jaime Moya
2003 Aspects of ASCI Code Verification and Validation. Albuquerque, NM: Sandia National Laboratories.
Trucano, Timothy, and Martin Pilch
2005 Measures of Modeling and Simulation as a Sandia DP Way-of-Life. Pp. 4. Albuquerque, NM: Sandia National Laboratories.
Trucano, Timothy, Martin Pilch, and William L. Oberkampf
2002 General Concepts of Experimental Validation of ASCI Code Applications. Pp. 137. Albuquerque, NM: Sandia National Laboratories.
Trucano, Timothy, et al.
2005 Software Engineering Intersections with Verification and Validation (V&V) of High Performance Computational Science Software: Some Observations. Pp. 15. Albuquerque, NM: Sandia National Laboratories.
Trucano, Timothy, and Laura Swiler
2004 Calibration, Validation, and Sensitivity Analysis: What's What and Who Cares? *In* Fourth International Conference on Sensitivity Analysis of Model Output. Pp. 25. Santa Fe, NM: Sandia National Laboratories.
Trucano, Timothy, Laura Swiler, and Takeru Igusa
2004 Calibration contra Validation: Characterization and Consequences. *In* Foundations for Verification, Validation and Accreditation in the 21st Century. Pp. 34. Phoenix, AZ: Sandia National Laboratories.
Trucano, Timothy, et al.
2006 Calibration, Validation, and Sensitivity Analysis: What's What. Reliability Engineering and System Safety:46.
Tsvetovat, Max, and Kathleen Carley
2003 Bouncing Back: Recovery mechanisms of covert networks. NAACSOS Conference 2003, Pittsburgh, PA, 2003. Vol. Day 3.
Tsvetovat, Max, Kathleen Carley, and K. Sycara
2001 Emergence of Market Segmentation: A Multi-Agent Model.
Tsvetovat, Max, Jeffrey Reminga, and Kathleen Carley
2003 DyNetML: Interchange Format for Rich Social Network Data. NAACSOS Conference 2003, Pittsburgh, PA, 2003. Vol. Day 2.
Tufarolo, John A., and Ernest H. Page
1996 Evolving the V&V; A process for the ALSP Joint Training Confederation. Coronado, California, United States: IEEE Computer Society.
Turbayne, Colin
1970 The Myth of Metaphor. Columbia, SC: University of South Carolina Press.
Turing, Alan
1950 Computing Machinery and Intelligence. Mind (236):433-460.
Turnley, Jessica Glicken
2004 Validation Issues in Computational Social Simulation. Pp. 8: Galisteo Consulting Group.
United States Department of Energy
1995 The Stockpile Stewardship and Management Program: Maintaining Confidence in the Safety and Reliability of the Enduring US Nuclear Weapon

Stockpile. O.o.D. Programs, ed. Washington, DC: United States Department of Energy.
United States Department of Energy
2004 Advanced Scientific Computing Research. Office of Science, ed. Pp. 79-87. Washington, DC: United States Department of Energy.
Urciuoli, Bonnie
2008 Skills and Selves in the New Workplace. American Ethnologist 35(2):211-228.
Van Asselt, Marjolein, and Jan Rotmans
2002 Uncertainty in Integrated Assessment Modeling: From Positivism to Pluralism. Climatic Change 54:75-105.
van Bloeman waanders, B. G. , et al.
2004 Sensitivity Technologies for Large Scale Sim, Sandia National Laboratories.
Van Brunt, Nicholas P.
1982 Simulator System for Logic Design Validation.
van den Broek, Egon L, et al.
2006 Formal Modeling and Analysis of Organizations. Lecture Notes in Artificial Intelligence 3913:18-34.
van Dijkum, Cor, Dorien DeTombe, and Etzel van Kuijk
1999 Validation of Simulation Models, Volume 403.
Van Horn, Richard L.
1971 Validation of Simulation Results. Management Science 17(5):247-258.
van Randwyk, Jamie, et al.
2005 Intrusion Detection and Monitoring for Wireless Networks. Pp. 158. Albuquerque, NM: Sandia National Laboratories.
Vicsek, Tamas
2002 Complexity: The bigger picture. Nature 418.
Watson, Donald
1984 Model, Metaphor and Paradigm. Journal of Architectural Education. 37(3/4).
Wauchope, R. Don
1992 Environmental Risk Assessment of Pesticides: Improving Simulation Model Credibility. Weed Technology 6(3):753-759.
Weick, Karl E., Kathleen Sutcliffe, and David Obstfeld
2005 Organizing and the Process of Sensemaking. Organization Science 16(4):409-421.
Werker, Claudia, and Thomas Brenner
2006 A practical guide to inference in simulation models.
Werner, Oswald
1972 Ethnoscience 1972. Annual Review of Anthropology 1: 271-308:37.

Whitner, R. B., and Osman Balci
1989 Guidelines for selecting and using simulation model verification techniques. Washington, D.C., United States: ACM.
Wikipedia
2007 Discrete event simulation.
Wilenski, Uri, and William Rand
2007 Making models match: Replicating an agent-based model. Journal of Artificial Societies and Social Simulation 10 (42).
Wilson, Robert V., et al.
2001 Comprehensive Approach to Verification and Validation of CFD Simulations—Part 2: Application for Rans Simulation of a Cargo/Container Ship. Journal of Fluids Engineering 123(4):803-810.
Windrum, Paul, Giorgio Fagiolo, and Alessio Moneta
2007 Empirical Validation of Agent-Based Models: Alternatives and Prospects. Journal of Aritificial Societies and Social Simulation 10(2).
Winsberg, Eric
1999 Sanctioning Models: The Epistemology of Simulation. Science in Context 12(2):1999.
Wise, John A., and David V Hopkin
1992 Verification and Validation of Complex Systems: Human Factors Issues. NATO Advanced Study Institute on Verification and Validation of Complex and Integrated Human-Machine Systems, Vimiero, Portugal, 1992. Springer.
Withers, R. F. J.
1959 Epistemology and Scientific Strategy. The British Journal for the Philosophy of Science 10(38): 89-102.
Wixon, Dennis, and John Whiteside
1985 Engineering for Usability: Lessons from the User Derived Interface. *In* CHI 1985: ACM.
Wolff, Gerard J.
2003 Neural Mechanisms for Information Compression by Multiple Alignment, Unification and Search: Cognition Research
Woods, David D., Leila Johannesen, and Scott S. Potter
1991 Human Interaction with Intelligent Systems: An Overview and Bibliography. SIGART Bulletin 2(5):39-50.
Wu, Fang, et al.
2004 Information Flow in Social Groups. Physica A:327-355.
Yahja, Alex
2006 Simulation Validation for Societal Systems Pittsburgh, PA: Carnegie Mellon University.
Yahja, Alex, and Kathleen Carley
2003 Generating Realistic Heterogeneous Agents: Computing Confidant-based Base Interaction Probabilities.
Yilmaz, Levent
2006 Validation and Verification of Social Processes within Agent Based Computational Organizational Models. Computational and Mathematical Organization Theory 12:283-312
Younger, Stephen M
1997 Confidence in the Absence of Full System Testing. Los Alamos, NM: Los Alamos National Laboratory.

Zeigler, Bernard P.
1984 Theory of Modeling and Simulation. New York: Krieger Publishing Co., Inc.
Zuk, Torre, et al.
2006 Heuristics for Information Visualization Evaluation. *In* BELIV 2006. Venice, Italy.

APPENDIX B: Memory & Reasoning Design Document

Design Document for the Neurobiology of Recollection: A Computational Model

1/01/2008

1 Abstract

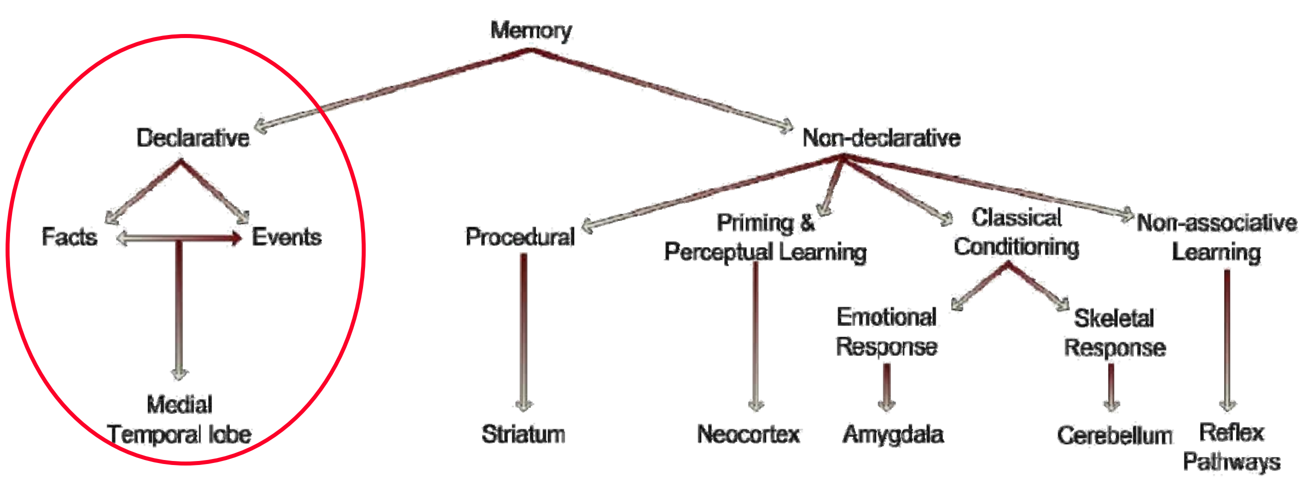
Our capacity for recollection is known to be supported by a system composed of several cortical association areas interacting with structures in the medial temporal lobe, and in particular, the hippocampus. There is a general consensus that the cortex is the repository of detailed representations of perceptions and thoughts and that the hippocampus supports the ability to bind together cortical representations and, when cued by part of a previous representation, to reactivate the full set of cortical representations that compose a recollective memory. Here we will propose a computational model of how the brain accomplishes retrospective recollection and prospective memory.

1.1 Review of Theories, Research, & Psychophysiological Models

Introduction

The overall objective of this project is to develop a computational model of how the hippocampal system might process information acquired during learning experiences leading to the consolidation of declarative memories. Below is the outline of the hypothesized structure of the anatomical circuitry of the hippocampal system, of the distinct functional roles of components of this system, and of how these system components might contribute to declarative processing and memory consolidation.

Like the Roman god Janus, memory looks both into the past and the future. Memory is usually thought of as a passive record of past events and acquired factual knowledge. But our adaptive application of memory is to make plans for our future actions. Therefore, our conscious lives are dominated by interactions between retrospective memory, the capacity for recollection of general knowledge and one's personal history of previous actions and their outcomes, and prospective memory, our intentional application of knowledge and history in directing ongoing decisions and behavior. Here we will propose a model of how the brain accomplishes retrospective recollection of memory. We will begin by outlining the experimental evidence on the cognitive and neural mechanisms of recollection, and then consider the interactions between retrospective and prospective memory from experimental studies in cognitive science. Then we will outline a formal model and its implementation in software.



What is recollection?

We have all been in the situation where we meet someone who seems highly familiar but we cannot recall who they are or why we know them. Sometimes, we just give up and say, “Don’t I know you?” Alternatively, when a clue or sufficient mental searching helps us retrieve a wealth of information all at once, including the name, where we met before, and the circumstances of the meeting. Considerable current research on recollection has focused the distinction between a vivid recollection, the lesser condition of a sense of familiarity with a particular person or object. Familiarity comes rapidly and reflects the strength match between a cue and a stored memory template. It is an isolated ability to identify a person or object as previously experienced. Recollection is typically slower and measured by the number of qualitative associations retrieved and the organization of the memory retrieved. Thus, recollections typically include not only the item sought in memory but also three other kinds of additional information: (1) the spatial and temporal context of the experience in which the item was previously encountered, (2) a replay of the sequence of events that compose an entire episode with that item, and (3) remembering additional related experiences with the item.

Furthermore, one brain area, the hippocampus, is critically involved in each of these aspects of recollection. Yonelinas et al. (2002) used ROC analysis on recognition memory performance to show that mild hypoxia that causes damage largely confined to the hippocampus resulted in a severe deficit in recollection but normal familiarity. A similar pattern of deficient recollection and preserved familiarity was reported in a patient with relatively selective hippocampal atrophy related to meningitis (Aggleton et al., 2005). Further consideration of the three properties of recollection introduced above provides insights into both the fundamental elements of recollection and the role of the hippocampus in memory processing.

Events are represented as items in the context in which they were experienced. A fundamental feature of recollection is memory for the spatial, temporal, and associational context in which experiences occur. There is a growing body of evidence that the hippocampus plays a critical role in remembering these contextual features and it does so by binding together representations of stimuli, actions, and places that compose discrete events. Functional imaging studies support the notion that the hippocampus is activated during the encoding or retrieval of associations among many elements of a memory, a characteristic of context-rich episodic memories (for review see Cohen et al., 1999; Eldridge et al., 2000; Maguire, 2001; Addis et al., 2004). For example, Henke et al., (1997) observed greater hippocampal activation when subjects associated a person with a house, as compared to making independent judgments about the person and house and others have found selective hippocampal activation during recollection of the context of learning in formal tests of memory (e.g. Davachi et al., 2003; Ranganath et al., 2003). The coding of associations extends beyond item and context associations such that the hippocampus is also selectively activated during the encoding or retrieval of verbal (Davachi & Wagner, 2002; Giovanello et al., 2003a) and face-name associations (Small et al., 2001; Zeineh et al., 2003; Sperling et al., 2003). Correspondingly, recent neuropsychological studies have found that recognition of associations is impaired even when recognition for single items is spared in amnesic patients (Giovanello et al., 2003, Turriziani et al., 2004). These studies reported impairment in recognition memory for associations between words or between faces or face-occupations pairs, as compared to normal performance in recognition of single items. At the same time, other functional imaging studies and characterizations of amnesia have suggested that the hippocampus is sometimes involved in both associative and single item recognition, highlighting the need to clarify the nature of associative information that composes an "event" (Squire et al., 2004). Nevertheless, these findings are generally consistent with the notion that the hippocampus plays a distinct role in recollection associated with binding features of items and their context to represent salient events (Eichenbaum et al., 2007).

Studies that employ animal models can provide compelling evidence on the effects of selective hippocampal damage. Several studies have shown that damage limited to the hippocampus results in deficits in forming a memory for the context or location where items were once experienced (reviewed in Mumby, 2001). In one recent study, rats were initially exposed to two objects in particular places in one of two environmental chambers (Mumby et al., 2002). In subsequent recognition testing, the place of the object or the context was changed. Normal rats increased their exploration of objects that were moved to new places or put in novel contexts. By contrast, rats with hippocampal damage failed to recognize objects when either the place or context was changed (see also Eacott & Norman, 2004).

Several investigators have argued that animals are indeed capable of remembering the temporal as well as spatial context in which they experienced specific stimuli (Clayton et al., 2003; Day et al., 2003). To further explore these aspects of episodic memory, we developed a task that assesses memory for events from a series of events that each involve the combination of an odor ("what"), the place in which it was experienced ("where"), and the order in which the presentations occurred ("when"; Ergorul & Eichenbaum, 2004). On each of a series of events, rats sampled an odor in a unique place along the periphery of a large open field. Then, memory for when those events occurred was tested by presenting a choice between an arbitrarily selected pair of odor cups in their original locations. Normal rats initially employed their memory of the places of presented cups and approached the location of the earlier experience. Then they confirmed the presence of the correct odor in that location. Animals with selective hippocampal damage fail on both aspects of this task even though their memory for independent features of location and odor items was intact. These findings indicate that the hippocampus is critical for effectively combining the "what", "when", and "where" qualities of each experience to compose the retrieved memory.

Neuro-anatomy studies

Studies on the firing properties of single neurons in animals provide insights into the nature of neural population representations in the hippocampus. Many studies have shown that hippocampal neurons encode an animal's location within its environment, and some view this as the principle function of hippocampal populations (Muller et al., 1999; Best et al., 2001). In addition, however, many other studies have shown that hippocampal neurons also fire associated with the ongoing behavior and the context of events as well as the animal's location (Eichenbaum et al., 1999). In the most direct examination of this issue, Wood et al (1999) directly compared spatial and non-spatial coding by hippocampal neurons by training animals to perform the same memory judgments at many locations in the environment. A large subset of hippocampal neurons fired only associated with a particular combination of the odor, the place where it was sampled, and the match-non-match status of the odor. In a similar study on the coding properties of hippocampal neurons in humans, Ekstrom et al. (2003) recorded in subjects as they played a taxi driver game, searching for passengers picked up and dropped off at various locations in a virtual reality town. They observed that many of these cells fired selectively associated with specific combinations of a place and the view of a particular scene or a particular goal. These and other studies indicate that, in rats, monkey, and humans, a prevalent property of hippocampal firing patterns involves the representation of unique associations of stimuli, their significance, specific behaviors, and the places where these events occur (see Eichenbaum et al., 2004).

Episodes are represented as sequences of events. We live our lives through personal experience, and our initial construction of reality within consciousness is a form of episodic buffer that contains a representation of the stream of events as they just occurred (Baddeley, 2000). In an early characterization of episodic recollection, Tulving (1983) distinguished episodic memory as organized in the temporal dimension, and contrasted this scheme with a conceptual organization of semantic memory. Tulving (1983) argued that the central organizing feature of episodic memory is that "one event precedes, co-occurs, or follows another". This is reminiscent of Aristotle's (350BC) characterization of vivid remembering: "Acts of recollection, as they occur in experience, are due to the fact that one thought has by nature another that succeeds it in regular order." These characterizations emphasize the temporal organization of episodic memories.

In humans memory for the order of events depends on hippocampal function. In a study using a design similar to that described above, Hopkins et al. (1995) found that patients with hypoxic brain injury involving in shrinkage of the hippocampus are impaired in memory for the order of a series of 6 words, pictures, or spatial locations. These patients were, however, also impaired in recognition of the items, undermining an unambiguous interpretation of a deficit in the order of the events independent of memory for the events. More recently, Spiers et al. (2001) reported a selective deficit in order memory independent of item memory in a patient with selective hippocampal damage due to perinatal transient anoxia (Vargha-Khadem et al., 1997). In this study the patient explored a virtual reality town in which he received objects from virtual characters. His recognition of the familiar objects was intact, but he was severely impaired in memory for the order in which he received objects, as well as for where he received them. Also, Downes et al. (2002) reported that patients with medial temporal lobe damage that included bilateral hippocampal damage were impaired in memory for the order of presentation of words for which recognition of the items was equivalent. Also, evidence from the deferred imitation task where subjects are required to remember an action sequence, indicate a critical role for the hippocampus (McDonough et al., 1995; Adlam et al., 2005). Thus, humans with hippocampal damage are impaired in memory for the order of events in unique episodes even in cases where recognition memory is intact.

Studies on animals also show that the representation of memories by the hippocampus incorporates not only items that must be remembered, but also the events that precede and follow. For example, Honey et al. (1998) provided a simple demonstration of the importance of temporal order in hippocampal processing, reporting that hippocampal lesions disrupted animals' normal orienting response when a pair of stimuli are presented in the opposite order of previous exposures. The specific role of the hippocampus in remembering the order of a series of events in unique experiences has been explored using a behavioral protocol that assesses memory for episodes composed of a unique sequence of olfactory stimuli (Fortin et al., 2002; see also Kesner et al., 2002). Memory for the sequential order of odor events was directly compared with recognition of the odors in the list independent of memory for their order. On each trial rats were presented with a series of five odors, selected randomly from a large pool of common household scents. Memory for each series was subsequently probed using a choice test where the animal was reinforced for selecting the earlier of two of the odors that had appeared in the series. In later sessions we also tested whether the rats could identify the odors in the list independent of their order, by was rewarding the selection of a novel odor against one that had appeared in the series. Normal rats performed both tasks well. Rats with hippocampal lesions could recognize items that had appeared in the series but were severely impaired in judging their sequential order.

How do hippocampal neuronal populations represent the sequences of events that compose distinct episodes? A common observation across many different behavioral protocols is that different hippocampal neurons become activated during every event that composes each

, including during simple behaviors such as foraging for food (e.g. Muller et al., 1987) as well as learning related behaviors directed at relevant stimuli that have to be remembered in studies that involve classical conditioning, discrimination learning, and non-matching or matching to sample tasks to tests and a variety of maze tasks (e.g. Hampson et al., 1993; for review, see Eichenbaum et al, 1999). In each of these paradigms, animals are repeatedly presented with specific stimuli and rewards, and execute appropriate cognitive judgments and conditioned behaviors. Corresponding to each of these regular events, many hippocampal cells show time-locked activations associated with each sequential event. Also, as described above, many of these cells show striking specificities corresponding to particular combinations of stimuli, behaviors, and the spatial location of the event. Thus, hippocampal population activity can be characterized as a sequence of firings representing the step-by-step events in each behavioral episode.

Furthermore, these sequential codings can be envisioned to represent the series of events and their places that compose a meaningful episode, and the information contained in these representations distinguishes related episodes that share common events and therefore could be confused. Recent studies on the spatial firing patterns of hippocampal neurons as animals traverse different routes that share overlapping locations provide compelling data consistent with this characterization. In one study, rats were trained on the classic spatial alternation task in a modified T-maze (Wood et al., 2000; see also Frank et al., 2000; Ferbinteanu and Shapiro (2003). Performance on this task requires that the animal distinguish left-turn and right-turn episodes that overlap for a common segment of the maze and requires the animal to remember the immediately preceding episode to guide the choice on the current trial, and in that way, the task is similar in demands to those of episodic memory. If hippocampal neurons encode each sequential behavioral event and its locus within one type of episode, then most cells should fire only when the rat is performing within either the left-turn or the right-turn type of episode. This should be particularly evident when the rat is on the “stem” of the maze, when the rat traverses the same set of locations on both types of trials. Indeed, a large proportion of cells that fired when the rat was on the maze stem fired differentially on left-turn versus right-turn trials. The majority of cells showed strong selectivity, some firing at over ten times the rate on one trial type, suggesting they were part of the representations of only one type of episode. Other cells fired substantially on both trial types, potentially providing a link between left-turn and right-turn representations by the common places traversed on both trial types.

Functional imaging studies in humans have also revealed hippocampal involvement in both spatial and non-spatial sequence representation. Several studies have shown that the hippocampus is active when people recall routes between specific start points and goals, but not when subjects merely follow a set of cues through space (Hartley et al., 2003). In addition, the hippocampus is selectively activated when people learn sequences of pictures (Kumaran & Maguire, 2006). Even greater hippocampal activation is observed when subjects must disambiguate picture sequences that overlap, parallel to the findings on hippocampal cells that disambiguate spatial sequences (Wood et al., 2000).

Memories are networked to support inferential memory expression. Further consideration of the cognitive properties of episodic memory suggest that related episodic representations might be integrated with one another to support semantic memory and the ability to generalize and make inferences from memories. Referring to how related memories are integrated with one another, William James (1890) emphasized that "…in mental terms, the more other facts a fact is associated with in the mind, the better possession of it our memory retains. Each of its associates becomes a hook to which it hangs, a means by which to fish it up by when sunk beneath the surface. Together they form a network of attachments by which it is woven into the entire tissue of our thought." James envisioned memory as a systematic organization of information wherein the usefulness of memories was determined by how well they are linked together.

There are two main outcomes of the linking of representations of specific experiences. One is a common base of associations that are not dependent on the episodic context in which the information was acquired. Thus when several experiences share considerable common information, the overlapping elements and common links among them will be reinforced, such that those items and associations become general regularities. The representation of these general regularities constitutes semantic "knowledge" that is not bound to the particular episode or context in which the information was encoded. The networking of episodic memories by common elements provides a mechanism for the commonly (albeit not universally, see Tulving, 2002) held view that semantic knowledge is derived from information repeated within and abstracted from episodic memories.

There is considerable evidence that hippocampal neurons indeed extract the common features among related episodes. In all the studies described above, a subset of hippocampal neurons encode features that are common among different experiences – these representations could provide links between distinct memories. For example, in the Wood et al. (1999) study on odor recognition memory, whereas some cells showed striking associative coding of odors, their match/non-match status, and places, other cells fired associated with one of those features across different trials. Some cells fired during a particular phase of the approach towards any stimulus cup. Others fired differentially as the rat sampled a particular odor, regardless of its location or match-non-match status. Other cells fired only when the rat sampled the odor at a particular place, regardless of the odor or its status. Yet other cells fired differentially associated with the match and nonmatch status of the odor, regardless of the odor or where it was sampled. Similarly, in

Ekstrom and colleagues' (2003) study on humans performing a virtual navigation task, whereas some hippocampal neurons fired associated with combinations of views, goals, and places, other cells fired when subjects viewed particular scenes, occupied particular locations, or had particular goals in findings passengers or locations for drop off. In studies that have recorded hippocampal neuronal activity as rats perform alternation tasks in a T-maze (Wood et al., 2000; Frank et al., 2000; Ferbintineau & Shapiro, 2003), whereas many cells distinguish overalapping actions and locations on the maze, some cells capture the common places and events between the different types of episodes.

The notion that hippocampal cells might reflect the linking of important features across experiences and the abstraction of common information was also highlighted in recent studies on monkeys and humans. Hampson et al. (2004) trained monkeys on matching to sample problems, then probed the nature of the representation of stimuli by recording from hippocampal cells when the animals were shown novel stimuli that shared features with the trained cues. They found many hippocampal neurons that encoded meaningful categories of stimulus features and appeared to employ these representations to recognize the same features across many situations. Kreiman et al., (2000a) characterized hippocampal firing patterns in humans during presentations of a variety of visual stimuli. They reported a substantial number of hippocampal neurons that fired when the subject viewed specific categories of material, e.g., faces, famous people, animals, scenes, houses, across many exemplars of each. A subsequent study showed that these neurons are activated when a subject simply imagines its optimal stimulus, supporting a role for hippocampal networks in recollection of specific memories (Krieman et al., 2000b). A subsequent study showed that some hippocampal neurons are activated a subject views any of a variety of different images of a particular person, suggesting these cells could link the recollection of many specific memories related to that person (Quiroga et al., 2005). This combination of findings across species provides compelling evidence for the notion that some hippocampal cells represent common features among the various episodes that could serve to link memories obtained in separate experiences.

The second outcome from a network of linked memories is a capacity to use the common elements to retrieve multiple memories that include that element. Furthermore, hippocampal representations could support a capacity to "surf" the network of linked memories and identify relationships and associations among items that were experienced in distinct memories and therefore are only indirectly related. A single cue could generate the retrieval of multiple episodic and semantic memories, and cortical areas can access these multiple memories to analyze the consequential, logical, spatial, and other abstract relationships among items that appeared separately in distinct memories. These logical operations on indirectly related memories can support inferences from memory. The activity of searching and surfing networks of memories, and then comparing and contrasting memories could underlie our awareness of memories and the experience of conscious recollection. The organization of linked experience-specific and experience-general memories with the capacity for association and inference among memories is called a "relational memory network."

In a series of studies, we have used a model system of rodent olfactory memory to explore the importance of the hippocampus in the linking memories and using the resulting relational networks to make associational and logical inferences from memory. One study examined the role of the hippocampus in making indirect associations between stimuli that were each directly associated with a common stimulus. Initially, normal rats and rats with hippocampal lesions were trained on a series of overlapping “paired associates” (Bunsey & Eichenbaum, 1996). On each trial, the rat was initially presented with one of two initial items in a pairing, and then had to select the arbitrarily assigned associate. For example, for training on the pairs A-B and X-Y, if A was the initial item, then the rat had to select B and not Y; conversely, if X was the initial item the rat had to select Y and not B. Then the rats were trained on a second paired associated list where the initial items were the second items in the first list and new items were the associates (B-C and Y-Z). Thus, when B was presented initially, the rat was required to select C and not Z; when Y was presented initially, the rats was then required to select Z and not C. After training on all four paired associates, the rats were tested on their knowledge of the indirect relations among the pairings. These tests involved presentations of an initial item from the first learned paired associates (A or X) followed by a choice between the second items of the later learned associates (C versus Z). Normal rats demonstrated their ability to express these indirect relations by selecting C when A was presented and Z when X was presented, whereas rats with selective hippocampal damage showed no capacity for this inference from memory. These findings, combined with observations on another transitive inference task (Dusek & Eichenbaum, 1997), indicate that the hippocampus is critical to binding distinct memories into a relational network that supports flexible memory expression.

In another experiment, rats learned a hierarchical series of overlapping odor choice judgments (e.g., A > B, B > C, C > D, D > E), then were probed on the relationship between indirectly related items (B > D ?). Normal rats learned the series and showed robust transitive inference on the probe tests. Rats with hippocampal damage also learned each of the initial premises but failed to show transitivity (Dusek & Eichenbaum, 1997). The combined findings from these studies show that rats with hippocampal damage can learn even complex associations, such as those embodied in the odor paired-associates and conditional discriminations. But, without a hippocampus, they do not interleave the distinct experiences according to their overlapping elements to form a relational network that supports inferential and flexible expression of their memories (see also Buckmaster et al., 2004).

Complementary evidence on the role of the hippocampus in networking of memories comes from two recent studies indicating that the hippocampus is selectively activated when humans make inferential memory judgments. In one study, subjects initially learned to associate each of two faces with a house and, separately, learned to associate pairs of faces (Preston & Gabrieli, 2004). Then, during brain scanning, the subjects were tested on their ability to judge whether two faces who were each associated with the same house were therefore indirectly associated with each other, and on whether they could remember trained face pairs. The hippocampus was selectively activated during performance of the inferential judgment about indirectly related faces as compared to during memory for trained face-house or face-face pairings. In the other study, subjects learned a series of choice judgments between pairs of visual patterns that contained overlapping elements, just as in the studies on rats and monkeys, and as a control they also learned a set of non-overlapping choice judgments (Heckers et al., 2004). The hippocampus was selectively activated during transitive judgments as compared to novel non-transitive judgments.

These findings indicate that the hippocampal relational network mediates the linking of distinct episodes that may contain items that have not been experienced in the same episode or in the same context. In doing so, the hippocampus plays a role in more than simply binding items within memories, but also mediates associations between distinct memories. During recollection, the hippocampus supports a capacity to generate multiple memories that share a common element, and the information contained within these memories can be used by many brain systems to make judgments about causal, logical, temporal, and spatial relations among the items in those memories (Cohen & Eichenbaum, 1993). Iterations of association, retrieval, and re-coding memories according to deduced relationships among the items would lead to the development of a systematic organization of items and episodes in memory wherein facts and events are linked to one another by a broad range of causal, logical, temporal, spatial, and other relevant relationships among the items. And this organization supports flexibility in the expression that is characteristic or recollective memory, specifically involving inferences between items that are only indirectly related.

How do the above described memory functions emerge from the circuitry of the hippocampus? The brain system that mediates retrospective and prospective memory is composed of several cortical association areas interacting with structures in the medial temporal lobe (MTL), and in particular, the hippocampus. There is a general consensus that areas of the cerebral cortex are specialized for distinct aspects of cognitive and perceptual processing that are essential to memory, and that the cortex is the repository of detailed representations of perceptions and thoughts. The MTL is the recipient of inputs from widespread areas of the cortex and supports the ability to bind together cortical representations such that, when cued by part of a previous representation, the MTL reactivates the full set of cortical representations that compose a retrospective memory. Areas of the cortex both direct the storage of memories in the MTL and interpret the reconstructed memories generated by the MTL to support prospective memory. This simple, anatomically based scheme provides the framework on which our model is built. In the following sections, we will describe in greater detail the functional components of this system and the pathways by which information flows among them, and a qualitative model of how they interact to support retrospective and prospective memory.

The anatomy of the brain system that supports memory is remarkably conserved across mammalian species (Manns & Eichenbuam, 2007). Information processing in this system occurs in three main stages. The first stage involves virtually every neocortical association area (Burwell et al., 1995; Suzuki, 1996). Each of these neocortical areas projects to one or more subdivisions of the parahippocampal region, which includes the perirhinal cortex, the parahippocampal cortex, and the entorhinal cortex. The subdivisions of the parahippocampal region are interconnected and send major efferents to multiple subdivisions of the hippocampus itself. Thus, the parahippocampal region serves as a convergence site for cortical input and mediates the distribution of cortical afferents to the hippocampus. Within the hippocampus, there are broadly divergent and convergent connections that could mediate a large network of associations (Amaral & Witter, 1989), and these connections support plasticity mechanisms that could participate in the rapid coding of novel conjunctions of information (Bliss & Collingridge, 1993). The outcomes of hippocampal processing are directed back to the parahippocampal region, and the outputs of that region are directed in turn back to the same areas of the cerebral cortex that were the source of inputs to the MTL.

Only highly pre-processed sensory information reaches the MTL, but these inputs come from virtually all higher-order cortical processing areas. Perhaps the most thoroughly studied cortical area afferent to the hippocampus is the inferotemporal (IT) cortex, the highest-order visual object processor in primates. Ablation of the inferotemporal cortex results in a visual-guided learning and deficits without impairment in visual fields, acuity, or threshold. The behavioral physiology of inferotemporal cortex is consistent with the data from ablation studies, showing that IT neurons are maximally driven by complex visual patterns, and the response properties of these cells are dependent on attentional mechanisms and reward association. Many IT neurons are preferentially responsive to a particular pattern, often one that is of obvious significance to the animal, including cells that respond selectively to faces. IT neurons respond differently to the same stimuli when they appear as stimuli to-be-remembered, or when they were novel versus familiar, and some cells maintain firing during the memory delay periods during performance of short term memory tasks. In humans, distinct ventral temporal areas that include and surround IT are activated by presentation of different categories of visual cues, including faces, tools, and animate objects (Martin, 2007; Kanwisher, 2007).

Other major inputs to the MTL arise from the posterior parietal area. Damage to this cortical area results in impairment in neglect of contralateral sensory stimulation across sensory modalities (Mountcastle et al., 1975; Andersen, 1989). One area within parietal cortex that has received particular interest is area 7a where most cells are visually driven. These cells have very large receptive fields and neuronal responsiveness is highly dependent on attentional factors. These cells respond best when the stimulus is the target of an eye or hand movement and they prefer moving stimuli but show little preference for stimulus form or color. These and other data indicate that the posterior parietal area is specialized for attention and egocentric spatial analyses including localization and visual and manual acquisition of targets in space. Also, areas of the parietal and temporal cortex are involved in complex perceptual processing essential to configuration of the conceptual contents of information that is the subject of recollection (e.g., Uncapher et al., 2006).

Additional major inputs to the MTL arise from several areas within the prefrontal cortex, a sensory-motor-limbic integration area involved in the highest-order cognitive functions including motor programming, vicarious trial and error, and memory (Fuster, 1995). In humans components of the prefrontal cortex mediate working memory, effortful retrieval, source monitoring, and other processing currently being specified that contribute critically to cognitive functions essential to recollection (Dobbins et al., 2002)). In addition, midline structures within the prefrontal and cingulate cortical areas have been identified as activated during processing of self-referential information that may be strongly related to autobiographical memory (Northoff & Bermpohl, 2004; Fink et al., 1996; Cabeza & St Jacques, 2007).

The nature of cortical inputs to the MTL differs considerably across mammalian species (Manns & Eichenbaum, 2006). The proportion of inputs derived from different sensory modalities also varies substantially between species, such that olfaction (e.g., rats), vision (e.g., primates), audition (e.g., bats), or somatosensation (e.g., moles) have become disproportionately represented in the brain in different animals (Krubitzer and Kaas, 2005). Nevertheless, the sources of information derived from prefrontal and midline cortical areas, as well as posterior sensory areas, are remarkably consistent across species.

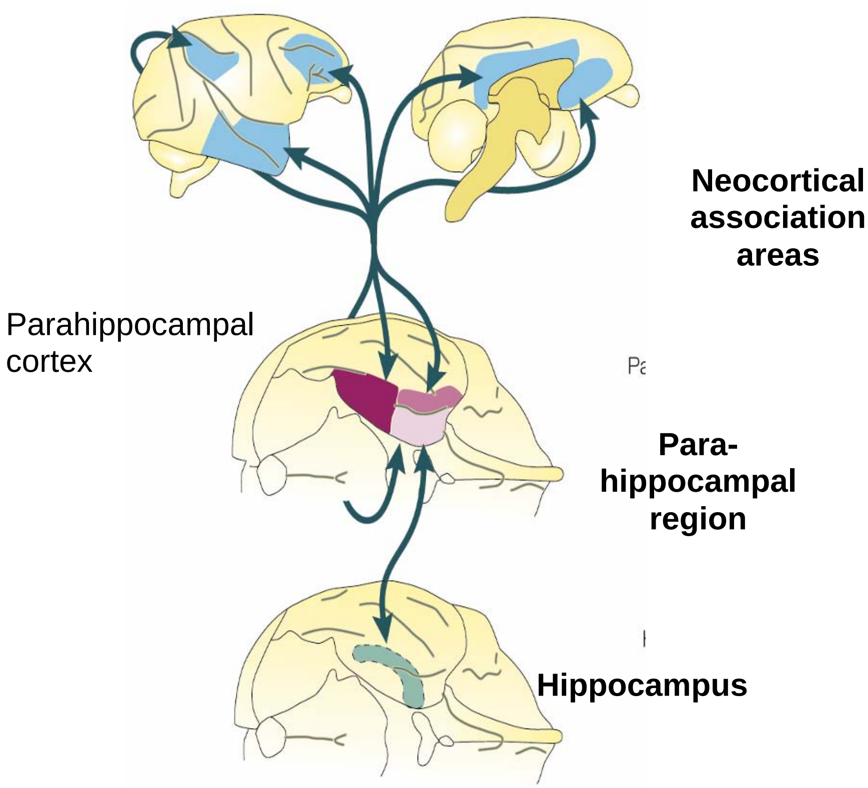
Despite major species differences in the neocortex, the organization of cortical inputs to the hippocampus is remarkably similar in rodents and primates. Across species, most of the neocortical input to the perirhinal cortex comes from association areas that process unimodal sensory information about qualities of objects (i.e., “what” information), whereas most of the neocortical input to the parahippocampal cortex comes from areas that process polymodal spatial (“where”) information (Suzuki & Amaral, 1994; Burwell et al., 1995. There are connections between the perirhinal cortex and parahippocampal cortex, but the “what” and “where” streams of processing remain largely segregated as the perirhinal cortex projects primarily to the lateral entorhinal area whereas the parahippocampal cortex projects mainly to the medial entorhinal area. Similarly, there are some connections between the entorhinal areas, but the “what” and “where” information streams mainly converge within the hippocampus. The cortical outputs of hippocampal processing involve feedback connections from the hippocampus successively back to the entorhinal cortex, then perirhinal and parahippocampal cortex, and finally, neocortical areas from which the inputs to the MTL originated.

2 Towards a model of a cortical-hippocampal memory system

Formally express the psychophysiological model. It is anticipated that the M&R team will need to develop this model from or building upon multiple sources in the literature, paying careful attention to acknowledge existing work on mathematical formalization. There should be formal descriptions of the inputs, outputs and state-content of the model, and a constructive (in the mathematical sense) description of how the model’s state and outputs are updated. The psychological specification of the model should read as a translation of the mathematical model into the language of psychology.

2.1 The functional organization of the cortical-hippocampal memory system

The anatomical evidence reviewed above suggests the following hypothesis about how information is encoded and retrieved during memory processing. During encoding, representations of distinct items (e.g., people, objects, events) are formed in the perirhinal cortex and lateral entorhinal area. These representations along with back projections to the “what” pathways of the neocortex can then support subsequent judgments of familiarity. In addition, during encoding, item information is combined with contextual (“where”) representations that are formed in the parahippocampal cortex and medial entorhinal area, and the hippocampus associates items and their context. When an item is subsequently presented as a memory cue, the hippocampus completes the full pattern and mediates a recovery of the contextual representation in the parahippocampal cortex and medial entorhinal area. Hippocampal processing may also recover specific item associates of the cue and reactivate those representations in the perirhinal cortex and lateral entorhinal area. The recovery of context and item associations constitutes the experience of retrospective recollection. In the succeeding sections, we will consider the evidence on the functional roles of these brain areas in support of this hypothesis.



Suzuki & Amaral, 1994;
Burwell & Amaral, 1998

Parahippocampal cortex and medial entorhinal area.

The cortical association areas perform considerable preliminary processing of stimuli, actions, thoughts, and events and can store this information briefly during processing. Recently perceived information will activate different cortical processing areas. These cortical areas will then create memory traces of the perceptual representations associated with this information. The perceptual representations are then matched with stored representations in the associated areas. This memory trace supports short-term recognition of this information that is associated with working memory. The information from the association areas of the different cortices are then processed by the MTL.

Within the MLT the parahippocampal region first receives the information. The parahippocampal region contributes to declarative memory by "buffering" specific representations that can be accessed and manipulated by the hippocampus; this processing can also support a sense of familiarity. Rats with damage to the parahippocampal region showed good retention at the shortest delay, but their performance declined abnormally rapidly across delays, showing a severe deficit within 1 minute. The hippocampus represents the critical relations among the items held by the parahippocampal region. There are two main routs by which the parahippocampal areas project into the hippocampal formation –the “long” and “short” routes (we will not make a distinction between the routes in this model). The parahippocampal region by itself mediates the representation of isolated items and can hold these representations in a memory buffer for periods of at least several minutes. This ‘intermediate-term’ memory function bridges the gap between the very brief period of immediate memory and the potentially permanent memory store.

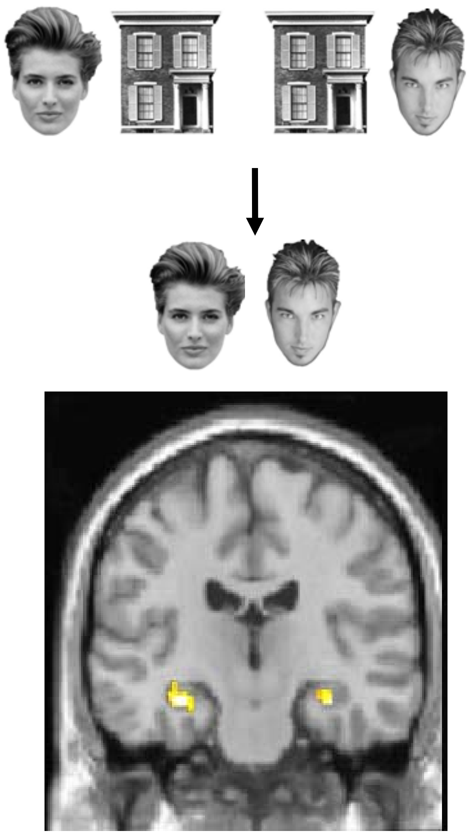
The parahippocampal cortex and medial entorhinal area may also be specialized for processing spatial context. Whereas perirhinal and lateral entorhinal neurons have poor spatial coding properties, parahippocampal and medial entorhinal neurons show strong spatial coding (Burwell and Hafeman, 2003; Hargreaves et al., 2005). Further, the immediate early gene fos is activated in perirhinal cortex by novel visual cues, but fos is activated in the postrhinal cortex by a spatial re-arrangement of the cues (Wan et al., 1999). In addition, whereas object recognition is impaired following perirhinal damage, object-location recognition is deficient following parahippocampal cortex damage in rats (Gaffan et al., 2004) and monkeys (Alvarado and Bachevalier, 2005). Similarly, perirhinal cortex damage results in greater impairment in memory for object pairings whereas parahippocampal cortex lesions results in greater impairment in memory for the context in which an object was presented (Norman and Eacott, 2005). Parallel findings from functional imaging studies in humans have dissociated object processing in perirhinal cortex from spatial processing in the parahippocampal cortex (Pihlajamaki et al., 2004). Furthermore, whereas perirhinal cortex is activated in association with the memory strength of specific stimuli (Henson et al., 2003), the parahippocampal cortex is activated during recall of spatial and non-spatial context (Ranganath et al., 2003; Bar and Aminoff, 2003).

Hippocampus.

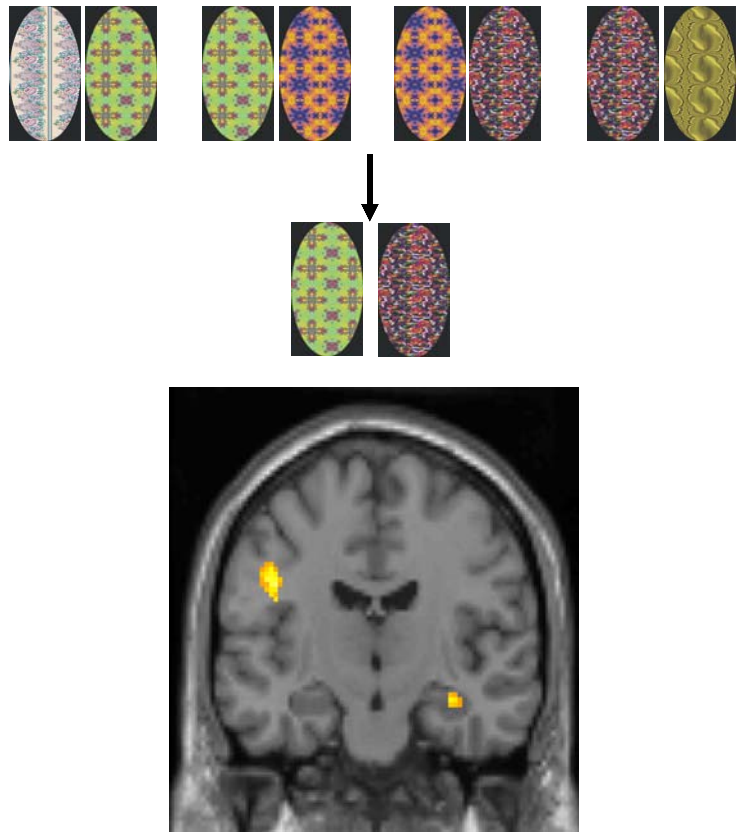
According to Eichenbaum et al., (1999), individual hippocampal cells encode regularities present in a person's every experience, including spatial and non spatial cues and behavioral actions. (215).

Compelling in support for differentiation of functions associated with recollection come from within-study dissociations that reveal activation of the perirhinal cortex selectively associated with familiarity and activity in the hippocampus as well as parahippocampal cortex was selectively associated with recollection (Deselaar et al., 2006; Davachi & Wagner, 2002; Davachi et al., 2003; Ranganath et al., 2003). These and many other results summarized in a recent review suggest a functional dissociation between the perirhinal cortex, where activation changes are consistently associated with familiarity, and the hippocampus and parahippocampal cortex, where activation changes are consistently associated with recollection (Eichenbaum et al., 2007). An outstanding question in these studies is whether the parahippocampal cortex and hippocampus play different roles in recollection. In particular, the above described findings on parahippocampal activation associated with viewing of spatial scenes suggests the possibility that this area is activated during recollection because recall involves retrieval of spatial contextual information. By contrast, the hippocampus may be activated associated with the combination of item and context information.

At the final stage of hippocampal system processing, the hippocampus is envoked, not to maintain a memory representation of single sensory cues or spatial information, but rather to process comparisons among the various current stimuli and between current stimuli and representations of previous stimuli, presumably those maintained at earlier levels of this system. Hippocampal processing appears to be quite different from the perceptual matching taking place in cortical and parahippocampal areas. Thus, hippocampal processing relies on cortical and parahippocampal inputs and presumably will exert its effects by modifying those inputs or by making connections among those cortical areas. In recognition memory, the hippocampus processes comparisons between current and previous stimuli as well as rich episodic and contextual information that goes beyond the strict perceptual properties on which cortical matchings are based; this may in some cases make a distinctive contribution to recognition memory. Thus, when the requirements of the task go beyond what can be accomplished by sensory matching processes supporting a sense of familiarity, requiring comparisons among experiences with items and their context and the flexible expression of memories, the entire hippocampal system contributes critically to a distinctly new capacity for declarative memory representation and this form of memory representation supports the capacity for conscious recollection.



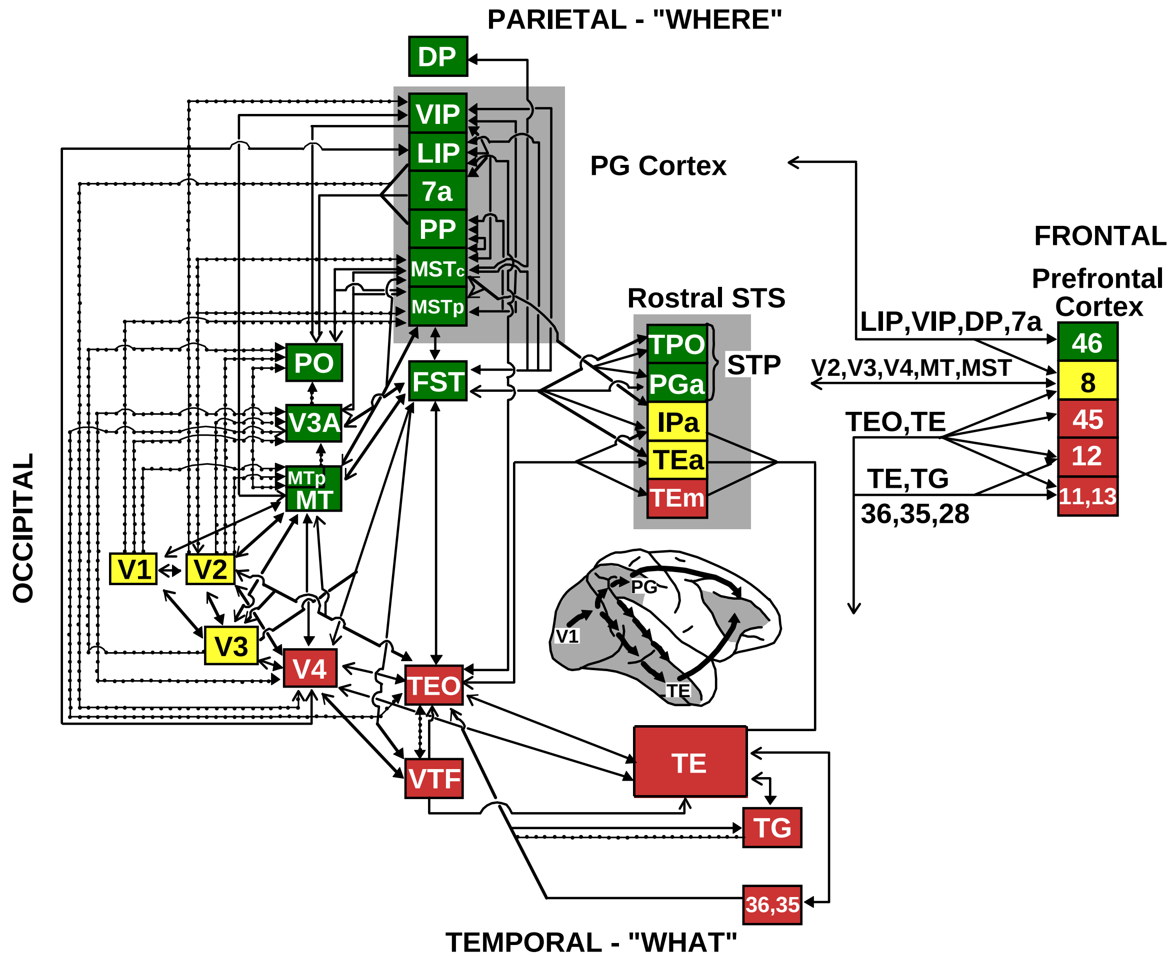
Preston et al., *Hippocampus* 2004



Heckers et al., *Hippocampus* 2004

The hippocampus has access to the much larger organization of item representations in cortical association areas via the parahippocampal region through accessing and re-accessing of information. Full relational memory organization comes about through multiple iterations of cortical input to the parahippocampal region and temporary storage there, followed by hippocampus-mediated relational processing that adds to or re-structures interconnections among parahippocampal and the cortical representations. The hippocampus itself mediates comparison and relating these individual representations to other memory representations, creating or modifying the overall memory organization according to the relevant relations between the items and the structure of any already-established memory organization that involves those items.

During encoding, representations of distinct items (e.g., people, objects, events) are formed in the perirhinal cortex and lateral entorhinal area. These representations along with back projections to the “what” pathways of the neocortex can then support subsequent judgments of familiarity. During encoding, item information is combined with contextual (“where”) representations that are formed in the parahippocampal cortex and medial entorhinal area, and the hippocampus associates items and their context. The hippocampus itself interacts with the neocortex only via the parahippocampal region, so one might expect that damage to the parahippocampus region would eliminate any relational processing contribution of the hippocampus. When an item is subsequently presented as a memory cue, the hippocampus completes the full pattern and mediates a recovery of the contextual representation in the parahippocampal cortex and medial entorhinal area. Hippocampal processing may also recover specific item associates of the cue and reactivate those representations in the perirhinal cortex and lateral entorhinal area. The recovery of context and item associations constitutes the experience of recollection. The combination processing functions of the parahippocampus and the hippocampus comprises declarative memory.



Hypothesis on how hippocampal system processing contributes to spatial representations.

Consistently, studies have shown that the hippocampus plays a special role in spatial representation (sss). A popular assertion for the last twenty years is that the hippocampus mediates the representation of physical space via a “cognitive map” of the environment.

The discovery of “place” cells within the hippocampus has bolstered this assertion of a cognitive map by providing a neural mechanism to encode spatial information (O’Keefe & Dostrovsky, 1971).

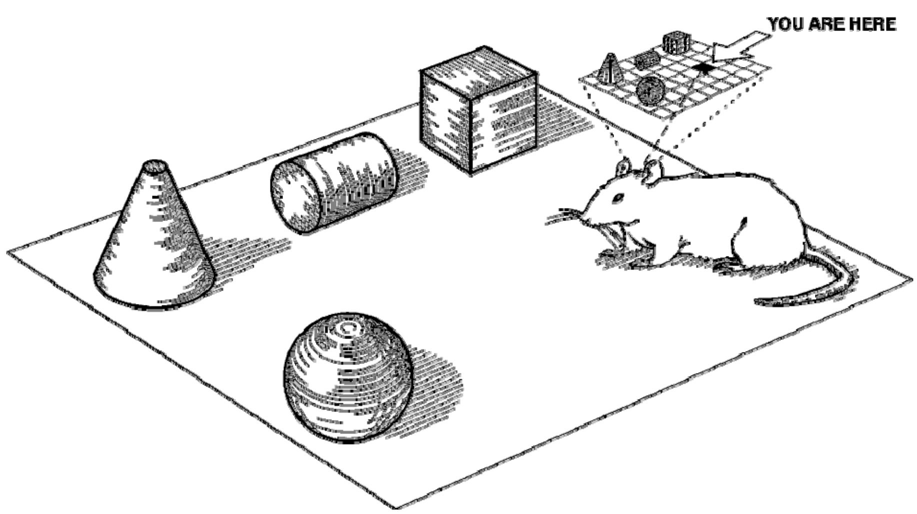
O’Keefe and Dostrovsky (1971) discovery of place cells. Found that some cells increased firing rate when a rat was at a particular location in its environment. In a large environment, one can correlate dramatic increases in a place cell’s firing rate when a rat arrives at a particular location, called the “place field.” Can go from a baseline of <1 spike/s, the firing rate can exceed 100Hz.

Place cells fire maximally when a person is in a fairly small, well-defined region of the environment. According to the cognitive-map theory, an environment is represented y a collection of place cells, each of which represents a specific region of space. The specific configuration of place cells provides an internal representation of the environment that affords an animal knowledge of its position relative to important locations. Distal visual cues, when present, appear to provide the preferred source of information used to support place-cell activity. However, distal cues can lose their control over place-cell activity if it is learned that the cues are unstable. Moreover, recent experience in an environment can exert a more powerful influence on place-field firing than presently available stable distal cues. In fact, changing the nature of the search strategy (to receive food, rewards, etc) from random to directed searching, caused a relocation of place fields in approximately 1/3 of the cells. In unchanging environments, place fields have been found to remain stable for over 153 days (Best et al., 2001).

The presentation of a familiar situation, even when some of the cues are missing or modified, results in retrieval of a previously established representation. However, when sensory cues are dramatically disrupted or the person moves to a new environment, the result is an almost complete reconfiguration of place codes. That is, the presentation of a larger number of changes in the environment outweighs the impact of any familiar features, results in a generation of a new representation. Experiments have shown that changing between two familiar environments without witnessing the switch is sufficient to induce generation of a new representation. However, when there is no prior experience, the initial exposure to the altered environment is not sufficient to generate a new spatial representation. This suggest that the combination of changes in environment shape and cue orientation is required to reach the threshold for pattern separation (Wilson et al., 2003).

Four distinct classes of place cells have been discovered; a) place cells that depend on location with respect to the distal cues; b) goal/landmark cells that fire close to the reward location or the landmarks, independent of their position in the environment; c) cells that fire only upon entering or leaving the start place, d) cells that are coded conjunctively or disjunctively over more than one of these dimensions.

Complementary to behavioral deficit findings showing that spatial information was encoded within the cellular activity of the hippocampal structures that are necessary for spatial learning and memory. Had the view that the hippocampus mediates a neural representation of physical space—i.e., a cognitive map (a holistic representation of space)

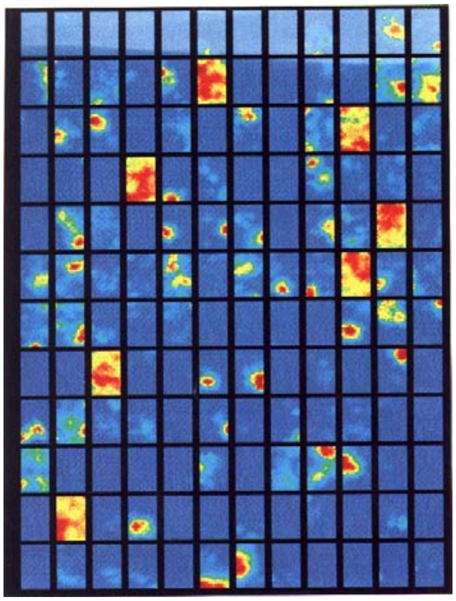


However, more recent electrophysiological studies have raised serious doubts about the notion of a cognitive map into question (Eichenbaum et at., 1999).

STUDIES FINDING GLOBAL MEMORY DEFICITS FOLLOWING HIPPOCAMPLE DAMAGE TO HUMANS AND OTHE RPRIMATES

For example, studies by Gothard and others (1996) have found that place cell activation is not bound to spatial representations. That is, a majority of the place cell firing were not associated with location-specific activity, but rather was associated with distances from a stimulus or goal. In fact, it has been argued by Morris (1990), that a major problem with the cognitive map idea is the lack of evidence demonstrating that place cells are organized as a spatial map or even that the primary function of the hippocampus is to encode spatial cues.

Place (pyramidal) cells in CA1of dorsal hippocampus



(Data from Wilson and McNaughton)

Results from recent studies of place cell firing patterns suggest that spatial representations are not cohesive, representing a global typology of the environment. Rather, the firings produce independent representations of spatial cues that are formed. These firing represent different subsets of spatial cues that are not bound to other spatial representations in the same environment (Eichenbaum, et al., 1999). For example, a study by Muller and Kubie (1987), found that a majority of cells become deactivated or change their place fields when the environment was altered. In cases involving goal-related landmarks, some cells fire relative to static environmental cues, while others fire relative to landmark-defined goal (Gothard et al., 1996). The Gathard at al., (1996) study found that most of the activated cells in the hippocampus were spatial to the extent that they fired a specific distances from specific stimuli and goals. Different specificities of patterns arise from the combination of input weights and the history of these inputs.

According to Eichenbaum (1999), place cells "exhibit movement-related firing patterns whenever particular movements are associated with meaningfully different events (p. 213).

Hypothesis regarding the encoding of spatial representations

With this in mind, it is proposed that initial experiences produce specific coding for particular conjunctions of stimuli, behaviors, and places that cooccur within the Hebbian time frame. Repeated repetitions of similar experiences will shape the responsiveness of the cells. Different types of pyramidal cells will have different firing tendencies. For example, broadly tuned cells will fire across a sequence of events. As these events are repeated in sequence, the strength of the associations is increased. In contrast, other cells will respond to novel events

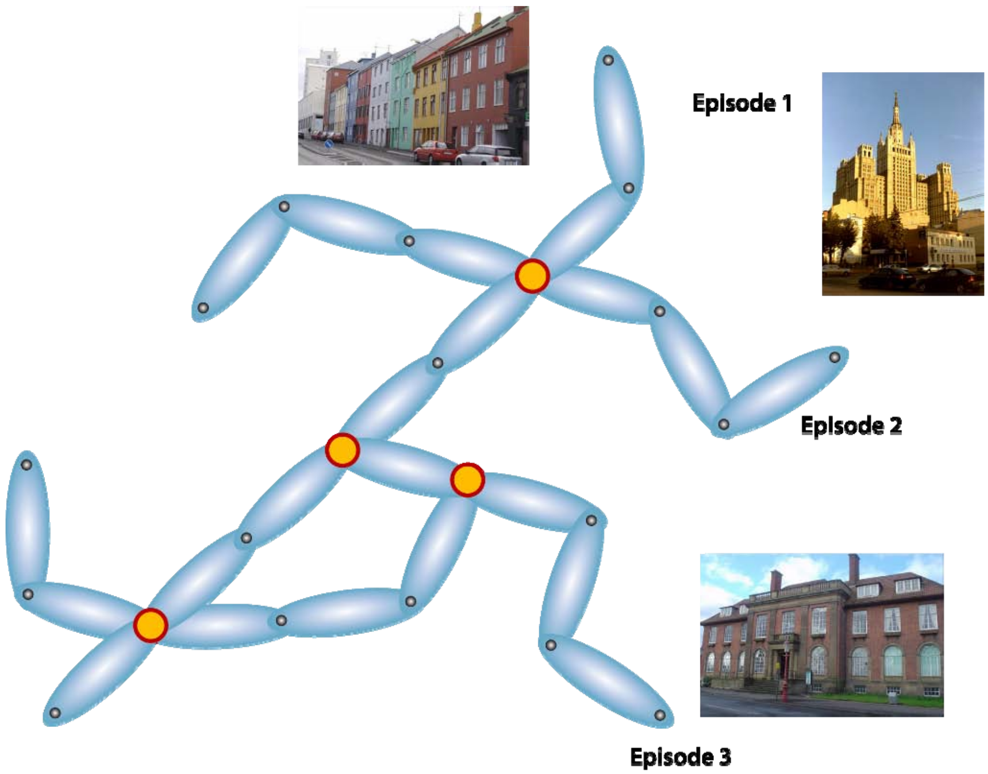
when particular combinations of salient inputs that influence a cell co-occur within 200ms (via Hebbian learning), the synaptic strengths of those inputs are altered to include this combination of inputs. This combination constitutes the representation of a specific "event."

Table 1. Properties of Hippocampal Neural Activity

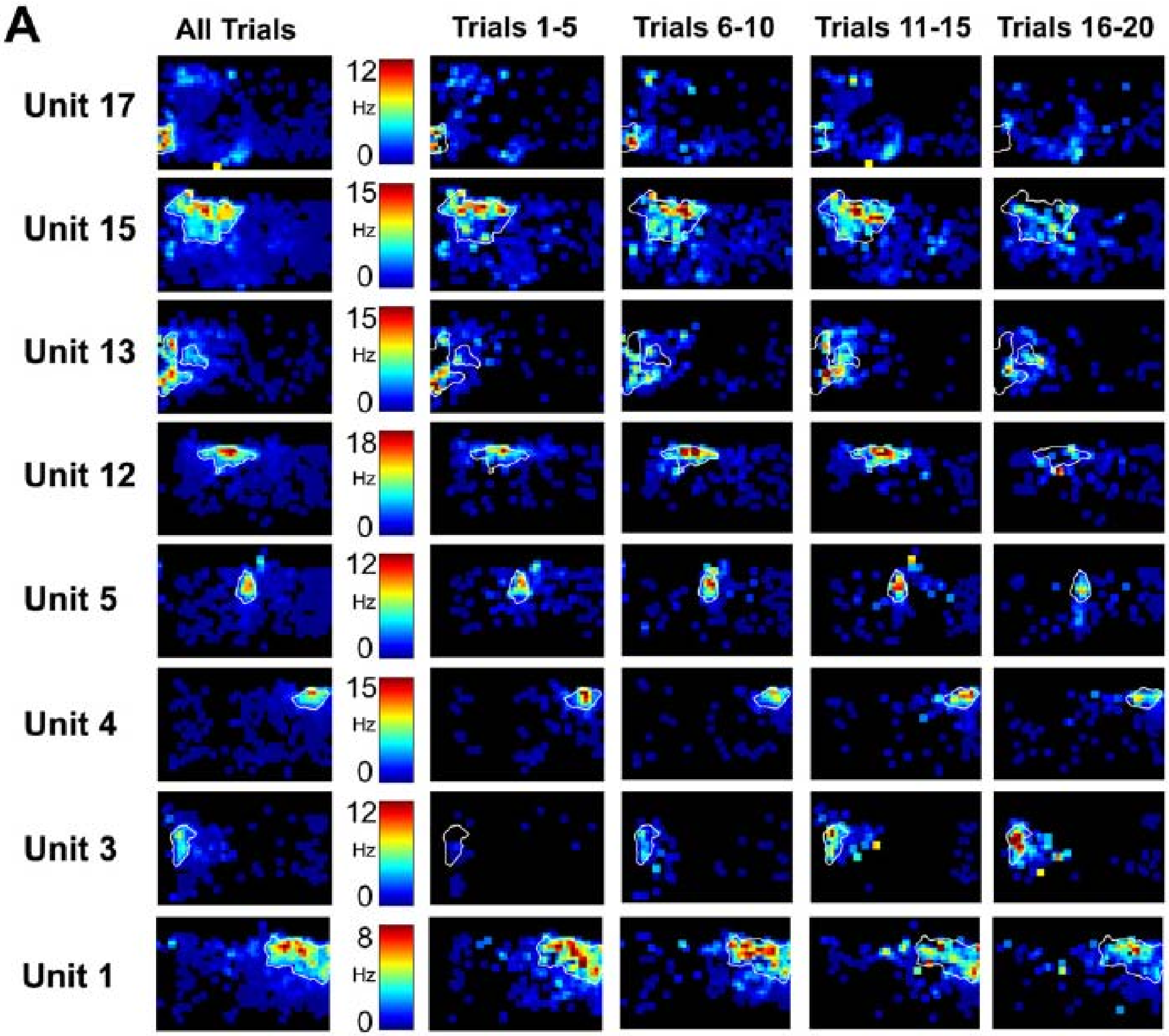
(1) The environment is not encoded in a continuous and systematic representation within the hippocampus. Instead, the codings of spatial and nonspatial features are organized in “clusters of neutrons that overrepresent some features of the environment at the expense of others.
(2) Hippocampal spatial firing patterns do not reflect the global topology of all the attended environmental cues. Instead, individual cells encode the relevant spatial relations among particular subsets of the cues.
(3) Hippocampal spatial firing patterns do not consistently represent the animal’s position among cues that compose an environment. Instead, the hippocampus creates distinct spatial representations, even for the identical spatial cues, under a vireity of conditions where the animal might consider itself undergoing different experiences within the same environment
(4) Within a broad variety of protocols in which animals learns regularities between stimuli, behavioral responses, and reinforcers, hippocampal neurons encode nonspatial stimuli and behaviors. These nonspatial firing correlates can be as robust and as prevalent as spatial firing patterns and, in a behavioral paradigm where distinctive events are distributed around the environment, they can be observed at all places where the associated events occur with regularity.
(5) The activity of many of the cells reflects the relevant spatial and nonspatial features of the task, whether or not the task is one that depends on hippocampal function.
(6) Both spatial and nonspatial representations are established very rapidly within the hippocampus
(7) Hippocampal neurons are activated during every phase of the performance of spatial and nonspatial tasks.
(8) Hippocampal neuronal activity reflects a broad spectrum of specificities. Some cells encode unique events, characterized by particular conjunctions of stimuli, behaviors, and the locations where these occur. Other cells represent sequences of events within behavioral episodes or specific features of events that are common across different behavioral episodes (Eichenbaum, et al., p. 216, 1999).

This representation may be generated region within minutes and can last for months. Blocking of LTP allows for plasticity of the memories (Austin et al., 1990).

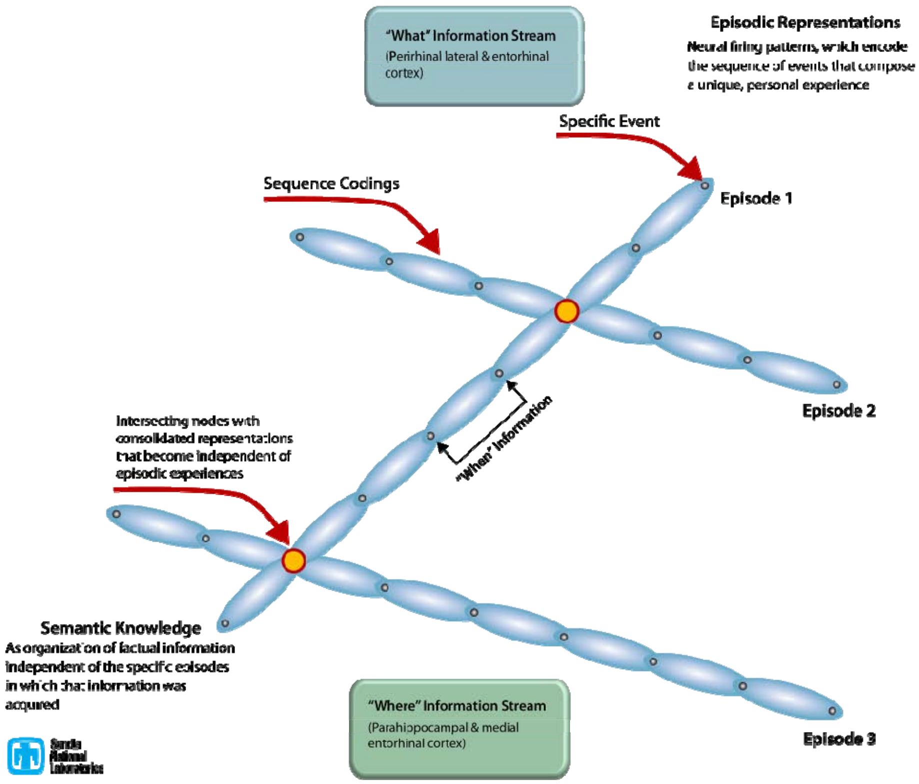
Data indicate that spatial firing patterns reflect independent representations of subsets of the spatial cues. Not bound to other spatial representations in the same environment
Spatial firing activity can be characterized as “spatial” only to the extent that they fire at specific distances from a particular stimulus or goal. From the data, it appears that place fields involve a collection of independent representations. Each one encoding the spatial relations between some subset of cues. Spatial representation are not bound as coordinates within a systematic framework for the global topology (i.e., no Cartesian “map”). Place cells exhibit movement-related firing patterns whenever particular movements are associated with meaningfully different events. In humans, a substantial number of hippocampal cells are activated to differentiate faces from objects, distinguish facial gender or expression, or distinguish new from familiar faces/objects.



Place cells exhibit movement-related firing patterns whenever particular movements are associated with meaningfully different events. Including spatial and nonspatial cues and behavioral actions. The conjunctive coding for a particular combination of stimulus and behavioral features constitutes the representation of a distinct “event.” Hippocampal representations form within minutes and are stable for months



Hypothesis on how hippocampal system processing contributes to memory consolidation. Once a set of hierarchical organizations is established and stabilized, it is difficult to smoothly add new items. This is not because a network cannot be altered to include new items be repetitive training, but because such novel training causes changes in an already-established network, resulting in turn, in catastrophic interference amount the already-existing items. A solution to this is to add a small network—a "hippocampus"—that could very rapidly acquire a representation of a new item and then have this small network slowly and gradually "train" the large network. In addition to the input from the hippocampus, the neocortex is repetitively exposed to old materials it was built to represent, thereby resulting in an interveaved learning regimen that intermixed repetitions of old and new representations. Eventually this process of interleaving produces an asymptotic state of the overall cortical representation, at which point it no longer benefited from hippocampal activations and thus no longer depend upon the hippocampus. Multiple iterations of the interactive processing described above may mediate the process of memory consolidation. Over extended time periods of weeks to months to years, new experiences that bear on the established organization reactivate established representations as well as add new ones, and these are processed together by this hippocampal circuit to weave the new information into the established relational network.



Because this network is so extensive and systematically interconnected, access to items via novel routes and in novel experiences is not only possible but also occurs continuously as we express memories to guide almost every aspect of daily life. These interactions, by feeding back and forth, can go on for a significant period, and may be reinstated repeatedly by experiences that bear partial similarity to the learning event. This repetitive processing could contribute to the "consolidation" of memories over very long periods. The period of consolidation is highly variable and may involve multiple stages. How might consolidation be accomplished by multiple iterations through the hippocampal system and how long does the process of consolidation take?

According to this model, the hippocampus contributes to consolidation via interplay with the cerebral cortex over a prolonged period (Eichenbaum et al., 1999). The sketch or indices that correspond to the cortical representations of a new episode are rapidly established in the hippocampus. Then, via explicit practice, repetition of the episode, or new related experiences, the hippocampus and cortex interact repetitively such that intracortical connections eventually support links between cortical representations. The present model specifically emphasizes the use of hippocampal elements to bind repeated or related experiences via common features.

Consolidation occurs in series of phases involving interactions within at least three levels of reciprocally connected areas and involves some distinctions between memories for specific events (episodic memories) and the abstraction of consistent, factual information from multiple events (semantic memory). Furthermore, this scheme emphasizes the hierarchical and reciprocal processing of information, initially through successive modality-specific cortical stages, then converging on the parahippocampal cortical areas surrounding the hippocampus, and finally reciprocal connections between the parahippocampal region and the hippocampus. The most detailed level of representations occurs in association areas, or perhaps in upstream modality specific-cortical processing areas. Then successively more abstract levels of representation occur in the parahippocampal region and then in the hippocampus.
Consolidation begins with interactions between the hippocampus and the parahippocampal region. For some period after learning, the episodic representations in the hippocampus serve to link convergent representations in the parahippocampal region. Feedback from the hippocampus is envisioned to mediate the development of episodic and semantic representations within the parahippocampal region by providing

an indirect pathway that drives the co-activation of parahippocampal neurons, enhancing the connections within their intracortical network. When links between representations have been acquired by parahippocampal cells, the memory can be considered to have consolidated there, in the sense that the memory abilities conferred by these cells would no longer require hippocampal feedback.

The next stage of consolidation involves a similar interplay between the cortical association areas and the parahippocampal region. Initially cortical associations are seen to depend on the parahippocampal region to supply linkages between their representations. By simultaneously driving cells in cortical areas and activating their intracortical connections, these linkages would be expected to mediate the ultimate development of links between the cortical association areas. The repeated invocation of hippocampal representations onto the cortex serves to re-organize cortical representations accommodating new information and new associations within the overall knowledge structure encoded there, and this structure contains both semantic knowledge and sequential information that composes long term episodic representations.

When this is accomplished the entire hippocampal circuit would no longer be necessary for the existence of permanent semantic and episodic representations. Thus, consolidation occurs in stages involving first a consolidation within the parahippocampal region and then later in the cortex— human amnesic patients with damage extending into the parahippocampal region have a more extended retrograde amnesia than those with selective hippocampal damage.

This model is consistent with, and its mechanisms are seen as mediating the interleaving of cortical memory representations proposed by McClelland et al. (Psych. Rev. 102: 419, 1995) to underlie consolidation. Thus, according to McClelland and colleagues, the hippocampus is proposed to be a device that rapidly stores new experiences, and for some period supports reinstatement of recent cortical memory representations. The cortex is viewed as mediating the consistent overall structure of knowledge and alters its representations via slow synaptic changes.

According to this model, the key feature of hippocampal-cortical interplay is that the hippocampus repetitively and intermittently reinstates episodic representations in the cortex and each reinstatement slowly interleaves new memories concurrently with reinstatements of previous memories that occur during everyday life. The result is an interleaving of new memories within the constantly evolving cortical representation of regularities.

This integrative processing, involving the interleaving of new representations among the existing structure, can benefit the cortical memory organization for a very long period. Indeed, contrary to the recent suggestion by Nadel and Moscovitch (Curr Opin. Neurobiol. 7: 217, 1997) that a prolonged consolidation period is not adaptive, according to the present view memory re-organization is seen precisely as an unending process.
On this view, the completion of consolidation is seen as a state at which integration of a new memory is asymptotic, that is, a state in which yet new experiences do not alter the relevant parts of the overall memory organization. When this state is achieved, removal of the hippocampus would not be expected to affect the operation of the cortical network. For some types of memory this might be achieved within days or weeks. Other memory experiences might benefit by integration with earlier formed memories over months or years. Thus the duration of consolidation is dependent on the nature of the learned material in terms of how many appropriate linkages across experience will benefit subsequent retrieval. To the extent that these are few and repeated frequently, consolidation will be completely readily. To the extent that memory for unique episodes benefits by linkage with many related episodes and facts, or continues to be reshaped by new experience, consolidation could go on for a lifetime.

Why might there be differences in the duration of memory consolidation for episodic and semantic memories? The current proposal addresses the distinction between episodic and semantic memory in both anterograde and retrograde amnesia. With regard to extended retrograde amnesia for episodic memories, for episodes that are only once experienced, it may be expected that few hippocampal-parahippocampal-cortical interplays occur. Consequently, the consolidation of cortical representations may require an extended period, or may never obtain sufficient co-activation to link the sequential representations of events. By this view, it is not unexpected that the retrieval of episodic memories may involve activation of the hippocampus for a prolonged period. For frequently re-experienced episodes, for rehearsed material, and when common features appear in repeated or related episodes, the rate of iterations of hippocampal and cortical processing is increased. Correspondingly, in these situations it may be expected that consolidation would be relatively rapid. Also, arousal-related modulatory processes that accelerate the cellular fixation of memories may also speed systems level consolidation, especially for highly emotional experiences.

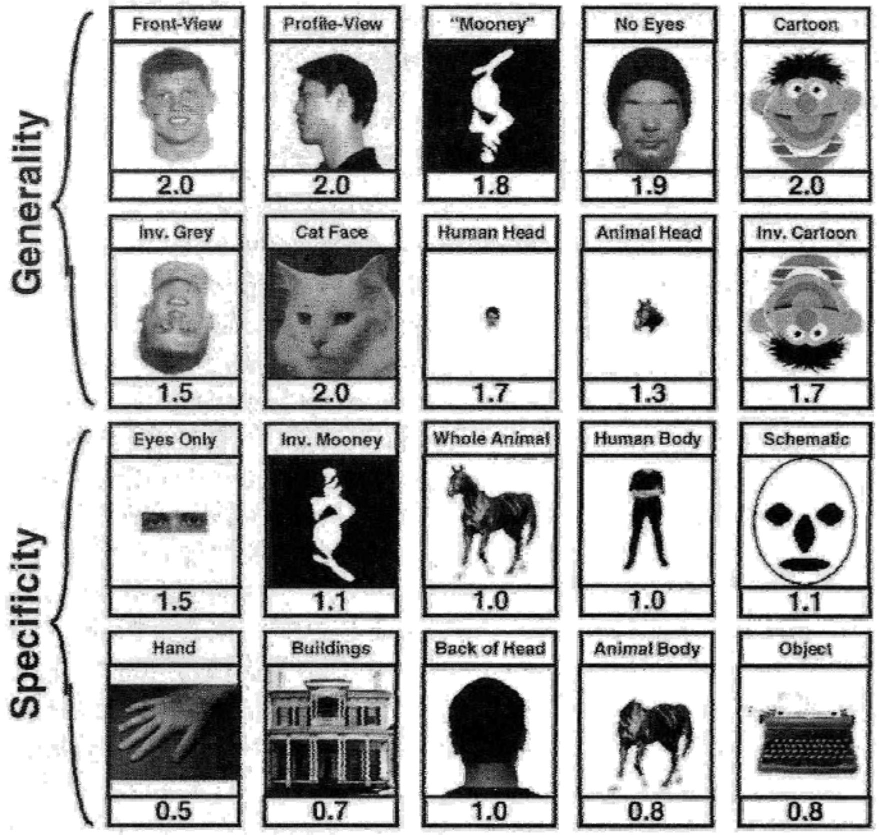
To the extent the particular information involves a single cortical area or overlapping cortical networks where intracortical links are already prevalent, it may be expected that the duration and extent of required hippocampal participation will be minimized. This view is consistent with Paller's (Memory 5:73, 1997) characterization of the hippocampus as providing coherence among representations mediated by distinctly separated cortical zones. He argued that semantic memories may be preserved in retrograde amnesia to the extent that they are mediated by a single cortical zone. Thus, it is expected that retrieval of some experiences that occur within a limited domain may involve activation of the hippocampus for a relatively brief period.

Neo Cortical Representation

The Conceptual Structure Account argues for a single amodal semantic store in which structure emerges from the distribution of features across categories. Amodal—arbitrary symbols within one semantic structure that stand for sensory-motor representations and for the environmental entities they represent. The Sensory/functional Theory argue that semantic representations are distributed across sensory and functional semantic processing regions of the brain that are closely linked to sensory and motor input/output processing channels. The domain-specific hypothesis argue that semantic representations are housed in processing channels specific to animals, plants, and nonliving objects that have evolved because of evolutionary pressures to avoid predators, etc.

Knowledge Types

Those who believe that semantic knowledge is instantiated across modality-specific processing regions that are situated beside and are closely linked to perceptual-processing areas. A concept's representation is the sum of the activation across primary sensory-processing channels, motor/action areas, higher order abstract-knowledge areas, and mediating association areas. Assume that semantic information related to specific input/output modalities is stored in regions close to their related sensory-input and motor-output processing regions. fMRI studies found that visual-processing areas are activated when accessing both visual and functional knowledge about living things but are activated when accessing only visual knowledge about nonliving things. Those who believe that semantic knowledge is transduced from patterns of activation in sensory regions into amodal semantic representations. Semantic knowledge is stored in a modality-neutral semantic processing area.



Distinguishing Among Concepts

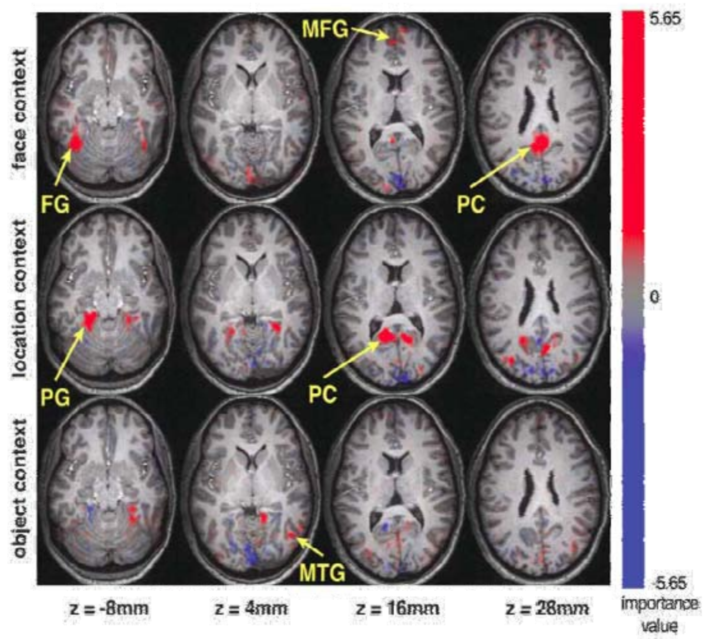
Correspond to the aspects of people's knowledge that discriminate among concepts in general or among sets of similar concepts. Similarity is related to the distinctiveness of a concept's features –the less distinctive the features, the more similar the concepts must be to one another and the closer they are in state space. Concept similarity is reflected in the proximity of the points that correspond to the concepts. Visual Complexity is the amount of object detail—can influence ability to integrate a picture's components into a coherent nameable object. Concept Familiarity increased experience with a concept's referents leads to a more robust representation. Correlated Features are said to be correlated if they tend to occur together in the same basic-level concepts.

Concepts based on structure

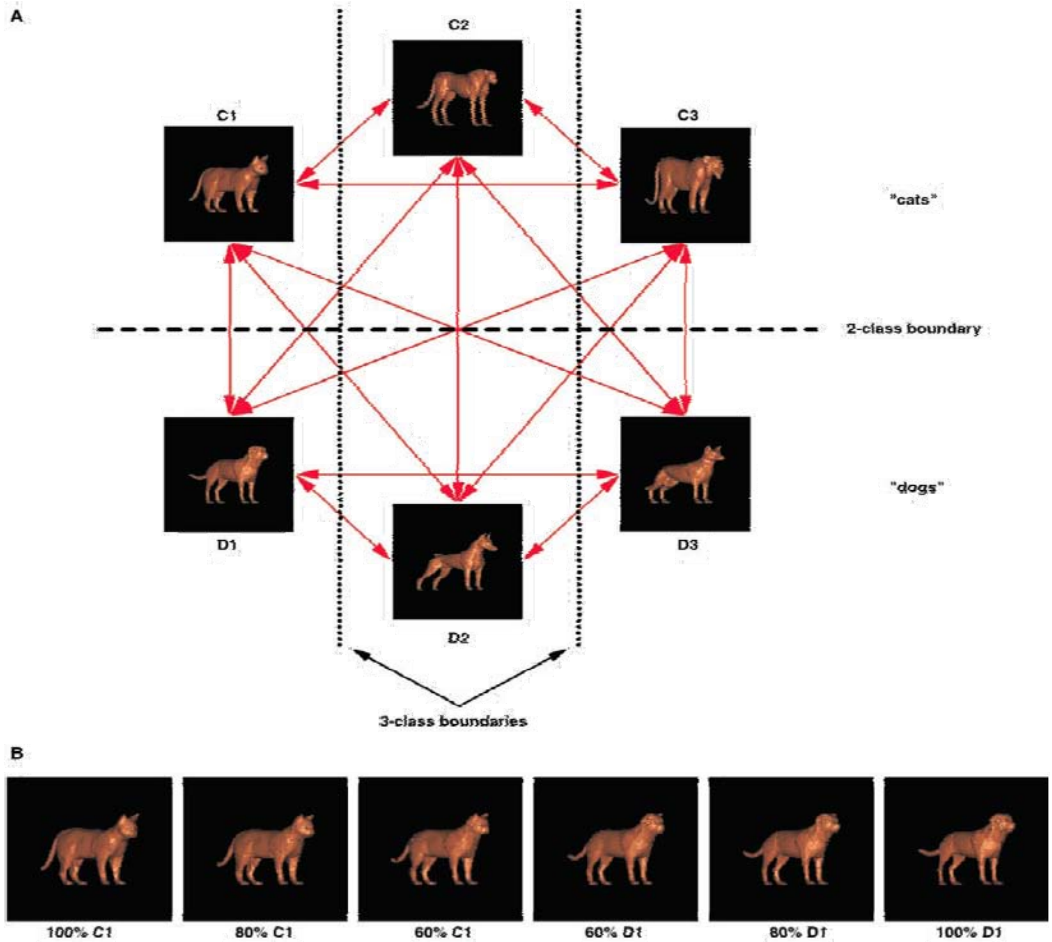
The distinction between nouns and verbs is universal. Current view that nouns and verbs map into ontologically distinct aspects of the environment (Gentner & Boroditsky, 1999). Idea that nouns refer to clusters of correlated properties that create chunks of perceptual experience. Some superordinate concepts are mass nouns (e.g., some furniture) and others are count nouns (e.g., an animal). Markman (1985) found that across languages, terms for categories at more abstract levels of a hierarchy are more likely to be mass nouns that are terms for categories at low levels of a hierarchy. Functional features are more important for artifacts, and features referring to internal structure are more important for natural kinds. Ahn found that the centrality of a feature to a category depends on the causal status of that feature relative to the other features in the category If a feature is thought to give rise to other features in the category, removing that casual feature affects category identity more than the removal of a noncausal feature does. Regardless of whether the category was a natural kind or an artifact, when functional features caused compositional features, functional features were considered more essential to category membership, whereas when compositional features caused functional features, compositional features were considered more essential to category membership. This suggest that the differences between artifact and natural kind categories may result from the fact that different kinds of features are causal in natural kinds are artifact categories.

Basic level vs subordinate and superordinate concepts

Rosch (1976) argued concepts form natural discontinuities that create a privileged level of categorization. However, basic level changes as a function of expertise. Learning may modify the constituent features or attributes of a concept (strongly supported by research). Different levels within an object hierarchy are useful for different kinds of tasks. Barsalou (1991) states that perceptual factors are more central than degree of informativeness to determine basic levels. Barsauou suggests that there may be a perceptual basic level, based primarily on shape and used largely during perception, and a more informational basic level, carrying more conceptual information and used for secondary categorizations during reasoning and communication. The typical privilege level correspond to genus level (e.g., blue jay, bass)



Polyn et al. Science. 310: 1963 2005



Freedman et al. 2001 Science 291: 312

Interrelated Concepts

There are several ways in which the characterization of concepts might be influenced by other concepts. Concepts are approximately equated with single words or phrases (e.g., dog). Concept learning involves learning to correctly categorize perceptual inputs into classes. A concepts characterization depends both on its representation and on the cognitive processes that operate on that representation. A concept is interrelated with respect to another concept to the extent that its characterization is influenced by the other concept

A concept might be characterized as a modified version of another previously acquired concept People's concepts typically contain both intrinsic and extrinsic features. Intrinsic features refer to parts or properties of the concept under scrutiny. Extrinsic features are represented as the relationship between two or more entities (will be influenced this way). Concepts compete to the extent that they are conceptually related neighbors. Rosch contends that natural categories are created so that there is as little overlap as possible between members of different categories. As one concept games control of a conceptual area, its competitor concepts lose control of the area (however, they can be positively interrelated if they are inductively or hierarchically related to each other.

A good diagnostic for locating isolated/interrelated concepts is asking the question "would this concept be used in this way if some/most/all other concepts were eliminated?" Barsalou (1993) have argued that concepts cannot simply gain their meaning from other concepts; concepts must also be grounded by perceptual, nonsymbolic properties. One way to conceive of an isolated concept is as a feature detector—i.e., does not need any information form other detectors, concepts, or theories in the system. Nondiagnostic features in categorization. One method for identifying isolated and interrelated conepts is to observe the influence of nondiagnostic features on categorization accuracy. Is a feature that does not, by itself, provide any information to choose between candidate categories. The difference between a diagnostic and a nondiagnostic feature is only relevant for interrelated concepts.

Exemplar Learning

Is central to the acquisition of many categories. Appears to be important to common taxonomic categories, such as apple, bird, relies on experiences with exemplars. It also serves to maintain accurate information about the kinds of entities in the world. – its physical properties, function, etc. The origins of categories determine their characteristics, so common taxonomic and goal-derived categories should differ. Common taxonomic categories exhibit prototype structure, with some exemplars being more typical than others. As people encounter a category's exemplars, they extract the exemplars' perceived characteristics and integrate them to from category knowledge. The representation that result from this learning can take the form of prototypes, exemplars, and/or definitions. Is automatic, bottom-up

Conceptual Combination learning, people derive new categories by manipulating existing knowledge in memory. Conceptual combination appears to be relatively active, top-down and effortful. Knowledge of categories may evolve through both exemplar learning and conceptual combination. People's intuitive theories about the world play central roles in the processing of exemplars, including the selection, interpretation, and integration of their perceived properties. As people extract perceptual characteristics from exemplars, the mechanisms of conceptual combination integrate this information with intuitive theories and other background knowledge to develop increasingly articulated accounts of the category.

People often acquire goal-derived categories through conceptual combination, in the absence of exemplars. Reasoning and conceptual combination during planning are central to acquiring categories. The formulation of goal-derived categories through conceptual combination in the absence of exemplars should preclude the abstraction of prototypical information.

Structure (cognitive representation of categories). Common taxonomic categories exhibit prototype structure with some exemplars being more typical of a category than others (e.g., robin vs chicken). As an exemplar becomes increasingly similar to prototypical information, it becomes increasingly typical. The ordering of exemplars according to typicality that results from similarity comparison constitutes the category's prototype structure. Prototype structure extends into the complement of the category, with non-member varying in how typical they are of the complement (e.g., butterfly, helicopter – members of non-birds). Prototype structure appears to be an implicit and emergent property that reflects the importance of prototypical information for a category, in conjunction with comparison and retrieval processes. This structure is central to the efficiency of classifying, production, and acquisition of exemplars, with typical exemplars being classified faster and more accurately than atypical exemplars. It is also central to reasoning about categories with typical exemplars facilitating syllogistic reasoning more than atypical exemplars.

Prototype Structure (are they different?)

Goal-derive categories exhibit the same prototype structure found in common taxonomic categories. When people judge typicality, the prototype structure that they produce for goal-derived categories are roughly as stable as those for common taxonomic categories. Even though a given goal-derived category may only occur to a few people on a few occasions, the causal principles that constrain it may be obvious and well known such that different people construct similar representations. Stability of Prototype Structure in both Taxonomic and Goal-derived Categories. Taxonomic and goal-derived categories are roughly equivalent in stability

Goal-derived Categories
Ideals (have to serve goal optimally)
» Whereas CT depends on exemplar learning, ideals do not
» Ideals arise from reasoning about categories with respect to goals
» Ideals exist in the representations of categories whose prototype structures they predict Frequency (number times perceived)
» Frequency does contribute to prototype structure via frequency of instantiation – not familiarity

Central tendency, ideals, and frequency causally determine prototype structure. For common taxonomic categories, CT dominated the prediction of typicality --all together account for 64% of variance. For goal-derived categories ideals and frequency account for 69% of variance.
• Thus, CT determines prototype structure in common taxonomic categories, but not in goal-derived categories.
• Exemplar learning is more central for common taxonomic categories that for goal-derived categories.

Because exemplar learning is necessary for acquiring CT, and because CT determines the prototype structure of common taxonomic categories, exemplar learning is central to the acquisition of these categories.

- Categories
- Goal-derived Categories

Ideals (have to serve goal optimally)

» Found that ideals are far more important than CT in determining the prototype structure of consumer categories, social categories, and names (e.g., cars, rock bands, etc).

Whereas exemplar learning produces CT information for common taxonomic categories, it does not produce CT information for goal-derived categories.

Evidence for Situated Simulation Theory

When a category is represented conceptually, the neural systems that processes it during perception and action become active in much the same way as if a category member were present. Sensory-motor representations in conceptual knowledge appear to fuse with incoming sensory information to construct perceptions.

It is hypothesized that participates simulate a category member to represent a category, and then consult the simulation to produce the requested information.
Damage to visual areas increases chances of losing "Living things" and damage to motor areas increases chances of losing "Manipulability artifacts."
Two level Theory of Conceptual Knowledge
At first level, reenactments of sensory-motor states are central to representing categories. The reenactment of sensory-motor states is central
At the second level, statistical representations in associated areas—much like those on the hidden layers of FF Nets—conjoin sensory-motor states into coherent representations
A primary function of statistical representations in association areas is to reactivate sensory-motor states in feature maps so that information relevant to the current task can be extracted from these simulations.
When it becomes necessary to process a category, its concept delivers a situated conceptualization that creates a multi-modal representation of what it would be like to process the category in that situation. A concept produces one of many possible representations that is tailored to the current context.

Together with the typically and membership data, the finding suggests that stable representations do not underlie concepts. For typically judgments, the average correlation between pairs of participants averaged around .40 across studies. According to the Situated action view, a concept is not a general description used over and over again across situations. Instead a concept is an ability or skill to construct specific representations that support different courses of situated action.

Psychophysiological Model Constraints

Provide a discussion of modeling assumptions, specifically a description of which domains the model applies to. For instance, is the model based upon research specific to high stress situations or research specific to text-based stimuli, etc?

Feasibility for Embodiment

Provide a discussion of what role the model can play in the cognitive framework of an embodied agent. This should include: how the model includes perception and action generation/control or interfaces to them; how the model's interface supports a host agent interacting with multiple entities; how spatial relationships impact inputs and outputs of the model.

Inputs

Describe the nature and structure of the input data. This would include an English description of the input data, data types and their ranges of values. The input can consist of structures that are composed from other structures.

Example: If this were a model of category formation, this section might specify how many dimensions the test data has, and what values each dimension can take (whether discrete or continuous). This section should also specify any constraints on the data such as mutually exclusive values along two different dimensions.

Outputs

Describe the structure of the output data. This would have the same form as the description of the inputs, i.e.: data types, ranges, etc. Following the category formation example above, the output might be a list of the inputs along with a category assignment for each one (say, category A or B). Alternately, if the input has a fixed sequence, the output could be a vector of A/B assignments without direct reference to the input values. Where necessary, this section should include English description of the output, especially if there are competing definitions for a term (like category assignment in this example).

Constructive procedure

List steps that describes exactly what to do with the input in order to get the outputs. Here is a reasonable place to mention internal states. The list of steps can take several forms. It may be one simple equation, or it may be pseudocode, or it may be a prosaic description of the procedure. Ideally, pseudocode would be structured using common CS conventions, but the actual text in each step would be written in full prose so that people from different specialties can understand it. Pseudocode could also take the form of a terse code-like manipulation followed by an extended comment that restates the same step in English. Diagrams may help illustrate the procedure.

Examples to illustrate how the model works

This could be a talk-through of one set of input data, showing how the output results

This template may be an oversimplification in some cases. For example, if a model included both a learning phase and a performance phase, then each one might have its own procedure. In that case, there may be two instances of the template, along with explanation of what state they share. Also, templates might be nested, with one template referring to a procedure that has its own template.

Embodiable Model Validation (if needed)
If a model from the literature was adapted or integrated with other models in order to obtain a model that is feasible for embodiment, then the resultant model (i.e., the candidate for us to implement) will need to be validated.

Psychophysiological Model Verification
Verify that the formal psychophysiological models express the actual phenomenon that is intended to be modeled. Generate predictions based on a mental walkthrough of the model against known experimental data (if available). Then compare the manually generated results against the reported results of the experiments. Explain any expected differences and justify expected correlations.

Review
Assuming that the model was developed by a subset of the team, a different subset should cross-check the model to ensure that it is consistent with the psychological literature.

Software Implementation of Model
Develop software that can execute/simulate the model.

Algorithmic Translation of Psychological Phenomenon
Fill out any implementation details not explicitly covered in the Model section. The combination of these details and the Model description should be sufficient to implement the model in an arbitrary computer language without further consultation with any of the authors of this document. It is acceptable, however, to assume that the implementer is an experienced software engineer who is familiar with standard algorithms and programming practices.

Record of Implementation
Give an account of how the software system was actually constructed. This is to be done according to coding conventions, documentation practices, and testing practices developed as much as possible prior to actual code-writing. Previously existing software may be integrated into this effort, provided that it already meets the agreed- upon standards or is adapted to meet them.

Documentation
The code will be written using literate programming practices. The comments in the code will include quotes of the Model so that it is clear how the two are connected. It should be possible to read the code or an extracted form of the documentation and get the same information that the Model section gives. (Note to non-programmers: There exist several tools, such as Doxygen, that can extract specially formatted comments from the code and output a document describing the software. This is far preferable to writing a separate document).

Verification

Verify that the Software models express the actual phenomenon associated with the psychological model. This will be accomplished by having the developers of the algorithmic models understand, review, and approve the Software models. After reviewers approve the Software models, they will formally 'sign-off' to this approval—see example below:

Reviewer 1	*Software Model 1*	*Approve Y* [ ] *N* [ ]
Reviewer 1	*Software Model 2*	*Approve Y* [ ] *N* [ ]
Reviewer 2	*Software Model 1*	*Approve Y* [ ] *N* [ ]

Full Experimental Validation

Once the design process is completed, it should be validated via a human subject experiment. The experiment(s) may incorporate one or more design processes. That is, software output from several design documents may be tested via a single experiment. The full validation experiment should be designed, at least initially, at the beginning of the psychological modeling phase. The manuscript should be written, to include experimental hypotheses before the experiment. The team should carefully document how input to the computer and input to the human subjects will be "comparable." The experiment and the documentation of the experiment should be formatted (e.g., introduction, method, results, discussion) according to APA guidelines. These experiments will typically be conducted at universities and run by academic consultants.

Experimental Validation Plan

Provide an experimental validation plan that will be executed at an appropriate time during the software development process. This plan will help guide the psychological model development as well as inform the software development. Multiple validations experiments may take place within one overall design process. Before each experimental iteration, an experimental plan should be discussed that includes what it is designed to accomplish and the method to do that. After each iteration, discuss the results of the experimentation.

References

Addis, D.R., Moscovitch, M., Crawley, A.P., and McAndrews, M.P. (2004) Recollective qualities modulate hippocampal activation during autobiographical memory retrieval. Hippocampus 14: 752-762.

Adlam, A.R., Vargha-Khadem, F., Mishkin, M., and de Haan, M. (2005) Deferred imitation of action dequences in developmental amnesia. Journal of Cognitive Neuroscience 17: 240-248.

Aggleton, J.P., Kyd, R.J., & Bilkey, D.K. (2004). When is the perirhinal cortex necessary for the performance of spatial memory tasks? *Neuroscience & Biobehavioral Reviews, 28*, 611-24.

Aggleton J.P., Vann S.D., Denby C., Dix S., Mayes A.R., et al. (2005) Sparing of the familiarity component of recognition memory in a patient with hippocampal pathology. *Neuropsychologia* 43(12):1810-1823.

Alvarado, M.C., & Bachevalier, J. (2005). Comparison of the effects of damage to the perirhinal and parahippocampal cortex on transverse patterning and location memory in rhesus macaques. *Journal of Neuroscience, 25,* 1599-1609.

Amaral, DG, Witter, MP. 1989. The three-dimensional organization of the hippocampal formation: A review of anatomical data. *Neuroscience* 31: 571-591

Amaral, D. G., & Witter, M. P. (1995). Hippocampal formation. In G. Pacinos (Ed.), *The Rat Nervous System,* (2^{nd} ed., pp. 443-493). San Diego, CA: Academic Press.

Andersen, R.A. (1989) Visual and eye movement functions of the posterior parietal cortex. Annual Review of Neuroscience 12:377-403.

Aristotle (350 BC). On Memory and Reminiscence. (J. I. Beare, Trans.).

Baddeley, A. (2000). The episodic buffer: A new component of working memory? Trends in Cognitive Science, 4, 417-423,

Bar, M. & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron, 38*, 347-58.

Best, P.J, White, A.M., and Minai, A. (2001) Spatial processing in the brain: the activity of hippocampal place cells. Annual Review of Neuroscience 24, 459-86.

Bliss, T. V. P., Collinridge, G. L. (1993). A synaptic model of memory: Long-term potentiation in the hippocampus. *Nature, 361,* 31-39.

Brown, M.W. & Xiang, J.Z. (1998). Recognition memory: Neuronal substrates of the judgment of prior occurrence. *Progress in Neurobiology, 55,* 149-89.

Brown, M.W. & Aggleton, J.P. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience, 2,* 51-61.

Buckmaster, C.A., Eichenbaum, H., Amaral, D.G., Suzuki, W.A. and Rapp, P. (2004) Enothrinal cortex lesions disrupt the relational organization of memory in monkeys. Journal of Neuroscience 24: 9811-9825.

Bunsey, M., & Eichenbaum, H. (1996). Conservation of hippocampal memory function in rats and humans. *Nature, 379*, 255-257.

Burwell, R.D., Witter, M.P. & Amaral, D.G. (1995). Perirhinal and postrhinal cortices of the rat: a review of the neuroanatomical literature and comparison with findings from the monkey brain. *Hippocampus, 5(5),* 390-408.

Burwell, R.D., & Hafemanm, D.M. (2003). Positional firing properties of postrhinal cortex neurons. *Neuroscience, 119,* 577-88.

Cabeza, R. and St. Jaques, P. (2007) Functional neuroimaging of autobriographical memory. Trends in Cognitive Sciences (in press).

Clayton, N. S., Bussey, T. J., & Dickinson, A. (2003). Can animals recall the past and plan for the future? *Nature Reviews Neuroscience, 4,* 685-691.

Cohen, N. J., Ryan, J., Hunt, C., Romine, L., Wszalek, T., Nash, C.(1999). Hippocampal system and declarative (relational) memory: Summarizing the data from functional neuroimaging studies. Hippocampus, 9: 83-98.

Cohen, N. J. & Eichenbaum, H. (1993). *Memory, Amnesia, and the Hippocampal System*. Cambridge, MA: M.I.T. Press.

Daselaar S.M., Fleck M..S, & Cabeza R. (2006) Triple dissociation in the medial temporal lobes: recollection, familiarity, and novelty. *Journal of Neurophysiology* 96:1902-1911.

Davachi, L. and Wagner, A. G . (2002). Hippocampal contributions to episodic encoding, Insights from relational and item-based learning. Journal of Neurophysiology, 88, 982-990.

Davachi, L., Mitchell, J. P. and Wagner, A. D. (2003). Multiple routes to memory, Distinct medial temporal lobe processes build item and source memories. Proceedings of the National Academy of Sciences. 100, 2157-2162.

Day, M., Langston, R., & Morris, R. G. M. (2003). Glutamate-receptor-mediated encoding and retrieval of paired-associate learning. *Nature*, *424*, 205-209.

Dobbins, I. G., Foley, H., Schacter, D. L., & Wagner, A. D. (2002). Executive control during episodic retrieval: Multiple prefrontal processes subserve source memory. *Neuron*, *35*, 989-996.

Downes, J. J., Mayes, A. R., MacDonald, C., & Humkin, N. M. (2002). Temporal order memory in patients with Korsakoff's syndrome and medial temporal amnesia. *Neuropsychologia, 40,* 853-861.

Dusek, J. A., & Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proceedings of the National Academy of Science, U.S.A., 94,* 7109-7114.

Eichenbaum, H., P. Dudchencko, E. Wood, M. Shapiro and H. Tanila (1999) The hippocampus, memory, and place cells: Is it spatial memory or a memory space? *Neuron* 23: 209-226

Eichenbaum H. 2004 Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron* 44:109-20.

Eichenbaum, H., Yonelinas A.R., and Ranganath, C. (2007) The medial temporal lobe and recognition memory. Annual Review of Neuroscience (In press).

Ekstrom, A.D., Kahana, M.J., Caplan, J.B., Fields, T.A., Isham, E.A., Newman, E.L., & Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature, 425*, 184-87.

Eldridge, L. L., Knowlton, B. J., Furmanski, C. S., Brookheimer, S. Y., & Engel, S. A. (2000). Remembering episodes: A selective role for the hippocampus during retrieval. Nature Neuroscience, 3, 1149-1152.

Ergorul, C., & Eichenbaum, H. (2004). The hippocampus and memory for “What”, “Where”, and “When”. *Learning and Memory, 11*, 397-405.

Ferbinteanu, J. & and Shapiro, M.L. (2003) Prospective and retrospective memory coding in the hippocampus. Neuron 40: 1227-1239..

Fink, G.R., Markowitsh, H.J., Reinkemeier, M., Bruckbauer, T., Kessler, J., and Heiss, W.-D. (1996) Cerebral representation of one’s own past: neural networks involved in autobiographical memory. Journal of Neuroscience 16: 4275-4282.

Fortin, N. J., Agster, K. L., & Eichenbaum, H. (2002). Critical Role of the Hippocampus in Memory for Sequences of Events. Nature Neuroscience, 5, 458-462.

Frank, L. M., Brown, E. N. & Wilson, M. (2000). Trajectory encoding in the hippocampus and entorhinal cortex. Neuron 27: 169-178.

Fuster, J. M. (1995). *Memory in the cerebral cortex.* Cambridge, MA: M.I.T. Press.

Gaffan, E.A., Healey, A.N., & Eacott, M.J. (2004). Objects and positions in visual scenes: effects of perirhinal and postrhinal cortex lesions in the rat. *Behavioral Neuroscience, 118*, 992-1010.

Gilbert, P.E., & Kesner, R.P. (2003). Localization of function within the dorsal hippocampus: the role of the CA3 subregion in paired-associate learning. *Behavioral Neuroscience, 117*, 1385-1394.

Giovanello, K.S., Verfaellie, M., and Keane, M.M. (2003) Disproportionate deficit in associative recognition relative to item recognition in global amnesia. Cognitive, Affective, and Behavioral Neuroscience 3: 186-194.

Giovanello, K.S., Schnyer, D.M., and Verfaellie, M. (2003) A critical role for the anterior hippocampus in relational memory: Evidence from an fMRI study comparing associative and item recognition. Hippocampus 14: 5-8.

Hampson, R. E., Heyser, C. J., and Deadwyler, S. A. (1993) Hippocampal cell firing correlates of delayed-match-to-sample performance in the rat. Behavioral Neuroscience, 107, 715-739.

Hampson, R.E., Pons, T.P., Stanford, T.R., Deadwyler, S.A. (2004) Categorization in the monkey hippocampus: a possible mechanism for encoding information into memory. Proceedings of the National Academy of Sciences 101:3184-3189.

Hargreaves, E.L., Rao, G., Lee, I., & Knierim, J.J. (2005). Major dissociation between medial and lateral entorhinal input to dorsal hippocampus. *Science, 308,* 1792-4.

Hartley, T., Maguire, E.A., Spiers, H.J., & Burgess, N. (2003). The well-worn route and the path less traveled: distinct neural bases of route following and wayfinding in humans. *Neuron, 37,* 877-888.

Heckers, S., Zalezak, M., Weiss, A.P., Ditman, T., and Titone, D. (2004) Hippocampal activation during transitive inference in humans. Hippocampus 14, 153-162.

Henson, R.N., Cansino, S., Herron, J.E., Robb, W.G., & Rugg, M.D. (2003). A familiarity signal in human anterior medial temporal cortex? *Hippocampus, 13,* 301-304.

Honey, R.C., Eatt, A. and Good, M. (1998) Hippocampal lesions disrupt an associative mismatch process. Journal of Neuroscience 18:2226-2230.

Hopkins, R. O. & Kesner, R. P. (1995). Item and order recognition memory in subjects with hypoxic brain injury. *Brain and Cognition, 27,* 180-201.

James, W. (1890). *The Principles of Psychology*. (1918 ed.). New York: Holt.

Kanwisher, N., Downing, P., Epstein, R., and Kourtzi, Z. (2007) Functional neuroimaging of visual recognition. In R. Cabeza & A. Kingstone, eds. Handbook of Functional Neuroimaging of Cognition. MIT Press: Cambridge. pp. 109-152.

Kesner, R. P., Gilbert, P. E., and Barua, L. A. (2002). The role of the hippocampus in memory for the temporal order of a sequence of odors. Behavioral Neuroscience, 116, 286-290.

Kesner, R.P., Hunsaker, M.R., & Gilbert, P.E. (2005). The role of CA1 in the acquisition of an object-trace-odor paired associate task. *Behavioral Neuroscience, 119,* 781-786.

Kreiman, K., Kock, C., and Fried, I. (2000a). Catgegory specific visual responses of single neurons in the human medial temporal lobe. Nature Neuroscience, 3, 946-953.

Kreiman, K., Kock, C., and Fried, I. (2000b). Imagery neurons in the human brain. Nature 408: 357-361.

Krubitzer, L., & Kaas, J. (2005). The evolution of the neocortex in mammals: how is phenotypic diversity generated? *Current Opinions in Neurobiology, 15,* 444-453.

Kumaran, D. & Maguire, E.A. (2006). The dynamics of hippocampal activation during encoding of overlapping sequences. *Neuron, 49,* 617-629.

Maguire E. A, (2001). Neuroimaging studies of autobiographical events memory. Philosophical Transactions of the Royal Society of London, Series B: 356, 1441-1452.

Manns JR and Eichenbaum H. (2006) Evolution of the hippocampus. In J.H. Kaas, ed. Evolution of Nervous Systems. Vol 3. Academic Press: Oxford pp. 465-490.

Martin, A. (2007) Functional neuroimaging of semantic memory. In R. Cabeza & A. Kingstone, eds. Handbook of Functional Neuroimaging of Cognition. MIT Press: Cambridge. pp. 153-186.

McDonough, L., Mandler, J.M., McKee, R.D., and Squire, L.R. (1995) The deferred imitation task as a nonverbal measure of declarative memory. Proceedings of the National Academy of Sciences 92: 7580-7584.

Mountcastle, V.B., J.C. Lynch, A. Georgopoulos (1975) Posterior partial association cortex of the monkey. Command functions for operations within personal space. Journal of Neurophysiology 38:871-908.

Muller, R.U., Kubie, J.L., and Ranck, J.B., Jr. (1987) Spatial firing patterns of hippocampal complex spike cells in a fixed environment. Journal of Neuroscience 7, 1935-1950.

Muller, R.U, Poucet, B., Fenton A.A., and Cressant, A. (1999) Is the hippocampus of the rat part of a specialized navigational system? Hippocampus. 9, 413-22.

Mumby, D.G., & Pinel, P.J. (1994). Rhinal cortex lesions and object recognition in rats. *Behavoral Neuroscience, 108,* 11-18.

Mumby, D.G. (2001) Perspectives on object recognition memory following hippocampal damage: lessons from studies on rats. Behavioural Brain Research 127, 159-181.

Mumby, D.G., Gaskin, S., Glenn, M.J., Scharamek, T.E. and Lehmann, H. (2002) Hippocmapal damage and exploratory preferences in rats: memory for objects, place, and contexts. Learning and Memory 9: 49-57.

Norman, G., & Eacott, M.J. (2005). Dissociable effects of lesions to the perirhinal cortex and the postrhinal cortex on memory for context and objects in rats. *Behavioral Neuroscience, 119,* 557-66.

Northoff, G. and Bermpohl, F. (2004) Corticla midline structures and the self. Trends in Cognitive Sciences 8:102-107.

Otto, T., & Eichenbaum, H. (1992). Complementary roles of orbital prefrontal cortex and the perirhinal-entorhinal cortices in an odor-guided delayed non-matching to sample task. Behavioral Neuroscience, 106, 763-776.

Pihlajamaki, M., Tanila, H., Kononen, M., Hanninen, A., Soininen, H., & Aronen, H.J. (2004). Visual presentation of novel objects and new spatial arrangements of objects differentially activates the medial temporal lobe areas in humans. *European Journal of Neuroscience, 19,* 1939-49.

Preston, A., Shrager, Y., Dudukovic, N.M. & Gabrieli, J.D.E. (2004) Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus, 14,* 148-152.

Quiroga R.Q., Reddy L., Kreiman G., Koch C., Fried I. (2005) Invariant visual representation by single neurons in the human brain. *Nature* 435:1102-1107.

Ranganath, C., Yonelinas, A.P., Cohen, M.X., Dy, C.J., Tom, S.M., and D'Esposito, M.D. (2003) Dissociable correlates of recollection and familiarity with the medial temporal lobes. Neuropsychologia 42: 2-13.

Suzuki, W.A., Zola-Morgan, S., Squire, L.R., & Amaral, D.G. (1993). Lesions of the perirhinal and parahippocampal cortices in the monkey produce long-lasting memory impairment in the visual and tactual modalities. *Journal of Neuroscience, 13,* 2430-51.

Suzuki W.A. & Amaral D.G. (1994) Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *Journal of Comparative Neurology 350,* 497-533.
Suzuki, W., & Eichenbaum, H. (2000). The neurophysiology of memory. *Annals of the NY Academy of Sciences, 911,* 175-91.

Yonelias, A.P., Kroll, N.E., Quamme, J.R., Lazzara, M.M., Sauve, M.J., Widaman, K.F., & Knight, R.T. (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nature Neuroscience, 5,* 1236-41.

Sperling, R., Chua, E., Cocchiarella, A., Rand-Giovannetti, E., Poldrack, R., Schacter, D.L, and Albert, M. (2003) Putting names to faces: Successful encoding of associative memories activates the anterior hippocampal formation. NeuroImage 20: 1400-1410.

Spiers, H. J., Burgess, N., Hartley, T., Vargha-Khadem, F., & O'Keefe, J. (2001). Bilateral hippocampal pathology impairs topographical and episodic memory but not visual pattern matching. *Hippocampus, 11,*715-725.

Suzuki W.A. & Amaral D.G. (1994) Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *Journal of Comparative Neurology 350,* 497-533.

Suzuki, W., & Eichenbaum, H. (2000). The neurophysiology of memory. *Annals of the NY Academy of Sciences, 911,* 175-91.

Tulving, E. (1983). *Elements of episodic memory*. Oxford: Clarendon Press.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53,* 1-25.

Turriziani, P., Fadda, L., Caltagirone, C., and Carlesimo, G.A. (2004) Recognition memory for single items and associations in amnesia patients. Neuropsychologia 42: 426-433.

Uncapher M.R., Otten L.J., Rugg M.D. (2006) Episodic encoding is more than the sum of its parts: an FMRI investigation of multifeatural contextual encoding. *Neuron* 52:547-556.

Vargha-Khadem, F., Gadin, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W. & Mishkin, M. (1997). Differential Effects of Early Hippocampal Pathology on Episodic and Semantic Memory. *Science, 277,* 376-80.

Wan, H., Aggleton, J.P., Brown, M.W. (1999). Different contributions of the hippocampus and perirhinal cortex to recognition memory. *Journal of Neuroscience*, *19,* 1142-48.

Wood E, Dudchenko PA, Eichenbaum H. (1999). The global record of memory in hippocampal neuronal activity. *Nature 397*: 613-16.

Wood E, Dudchenko P, Robitsek JR, Eichenbaum H. (2000). Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron, 27,* 623-33.

Zeineh, M.M., Engel, S.A., Thompson, P.M., and Brookheimer, S.Y. (2003) Dynamics of the hippocampus during encoding and retrieval of face-name pairs. Science 299: 577-580.

Internal Distribution

1	MS 0370	George Backus, 1433
1	MS 0370	Laura A. McNamara, 1433
1	MS 1316	Scott A. Mitchell, 1415
1	MS 1322	Alexander Slepoy, 1435
1	MS 0370	Timothy G. Trucano, 1411
1	MS 0899	Technical Library, 9536 (electronic copy)
1	MS 0123	Donna L. Chavez, 1011

Sandia National Laboratories