SANDIA REPORT
SAND2008-7968
Unlimited Release
Printed December 2008

A Surety Engineering Framework to Reduce Cognitive Systems Risks

David E. Peercy, Wendy L. Shaneyfelt, Eva O. Caldera, and Thomas P. Caudell

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.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

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

SAND2008-7968
Unclassified Unlimited Release
Printed December 2008

A Surety Engineering Framework to Reduce Cognitive Systems Risks

David E. Peercy
Weapon System and Software Quality
Wendy L. Shaneyfelt
Cognitive System Research and Applications
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185

Eva O. Caldera
Associate Director, Institute for Ethics
Thomas P. Caudell
Director, Center for High Performance Computing Visualization Laboratory
University of New Mexico
1313 Goddard SE,
Albuquerque, NM 87106

Abstract

Cognitive science research investigates the advancement of human cognition and neuroscience capabilities. Addressing risks associated with these advancements can counter potential program failures, legal and ethical issues, constraints to scientific research, and product vulnerabilities. Survey results, focus group discussions, cognitive science experts, and surety researchers concur technical risks exist that could impact cognitive science research in areas such as medicine, privacy, human enhancement, law and policy, military applications, and national security (SAND2006-6895).

This SAND report documents a surety engineering framework and a process for identifying cognitive system technical, ethical, legal and societal risks and applying appropriate surety methods to reduce such risks. The framework consists of several models: Specification, Design, Evaluation, Risk, and Maturity. Two detailed case studies are included to illustrate the use of the process and framework. Several Appendices provide detailed information on existing cognitive system architectures; ethical, legal, and societal risk research; surety methods and technologies; and educing information research with a case study vignette. The process and framework provide a model for how cognitive systems research and full-scale product development can apply surety engineering to reduce perceived and actual risks.

ACKNOWLEDGEMENTS

The authors for this report include personnel from Sandia National Laboratories and UNM who have expertise in cognitive science, surety technologies, and legal and ethical issues. In addition, there were participants from the Potomac Institute for Policy Studies, Department of Homeland Security, and other areas within Sandia National Laboratories who provided their reviews and suggestions on the approach and results of this effort.

The authors would like to thank the following Sandia National Laboratories personnel who have added discussions, reviews, research results, and in some cases extensive time to improve the information in this report. However, the content of this report is the sole responsibility of the authors and does not represent approval of the results by these persons.

Kevin Dixon, 06341

Michael Bernard, 06341

Chris Forsythe, 06341

Wendell Jones, 00500

Jonathan McClain, 06343

Laura McNamara, 01433

John Russell, 06414

Richard Sarfaty, 12336

Gerry Sleefe, 06140

Ann Speed, 06343

Timothy Trucano, 01411

John Wagner, 06141

Gerry Yonas, 00700

In addition, there have been numerous interactions with personnel outside Sandia concerning this effort. Apologies to those who may have been left off the list, but some of these persons include:

Potomac Institute for Policy Studies

Jim Giordano, Alexis Jeannotte, Dennis McBride, Leah Reeves, Kathryn Schiller

Department of Homeland Security

Ji Sun Lee, Richard Lempert, Sharla Rausch

Mitre Corporation

Brad Minnery

Table of Contents

LIST OF FIGURES

List of Tables

This page blank except for this statement.

EXECUTIVE SUMMARY

This report describes the results of the FY08 Lab Directed Research and Development (LDRD) project, #126630 titled: "Investigating Frameworks for Application of Surety Methods to Reduce Development and Operational Risks of Cognitive Sciences and Technologies." The purpose of this LDRD was to identify a process and framework for the application of surety methods to reduce the development and operational risks associated with cognitive systems. A spectrum of such risks was identified in an earlier LDRD project #105306, Investigating Surety Methodologies for Cognitive Systems.

Technical Approach

Sandia National Laboratories teamed with the University of New Mexico to further investigate how the previously identified spectrum of cognitive system risks might be reduced. The technical approach was to:

(1) investigate existing frameworks for cognitive systems as well as potential examples of existing and/or futuristic cognitive systems for which this research might apply;

(2) derive a general framework for modeling the specification, design architecture, evaluation for verification and validation, and quality/risk indicators of cognitive systems;

(3) incorporate requirements within the specification model that address principles of ethics, legal, societal, and surety that address concerns identified in the cognitive system risk spectrum;

(4) incorporate existing cognitive system designs within the cognitive system architecture model;

(5) incorporate surety engineering methods such as safety, reliability, and security within the evaluation and risk models to address potential science-based vulnerabilities in the development and/or operational use of cognitive systems;

(6) apply a risk-informed decision process that can be applied to manage and hopefully reduce the risk of identified cognitive system vulnerabilities; and

(7) develop a case study that illustrates the value of and how to apply the derived general framework and risk-informed decision process.

Cognitive System definition:

> *Cognitive systems are implementations of technologies that utilize as an essential component(s) one or more plausible models of human cognitive processes.*

Cognitive systems have one or more functions that model a human's cognitive tasks. Such systems may be implemented as a computational system, biological system, or some combination. Cognitive systems include implementations that advance/augment human cognition, simulate human cognitive tasks either for understanding or operational use, or perhaps limit/reduce human cognition capabilities.

Results

Quality is the result of managing vulnerabilities to a targeted risk. Cognitive systems, by the very nature of their applications, must achieve a reasonable level of quality. A generic systems/surety engineering framework architecture has been developed that would also apply to any system – but in this report is being applied to cognitive systems. Four models are part of the architecture: Specification Model, Design Model, Evaluation Model, and Risk Model, as illustrated in the Figure EX-1 below. In addition, a Maturity Model is applied appropriate to the life cycle activities to represent the plausibility of the cognitive system implementation with regard to identified risks as part of the risk-informed decision process.

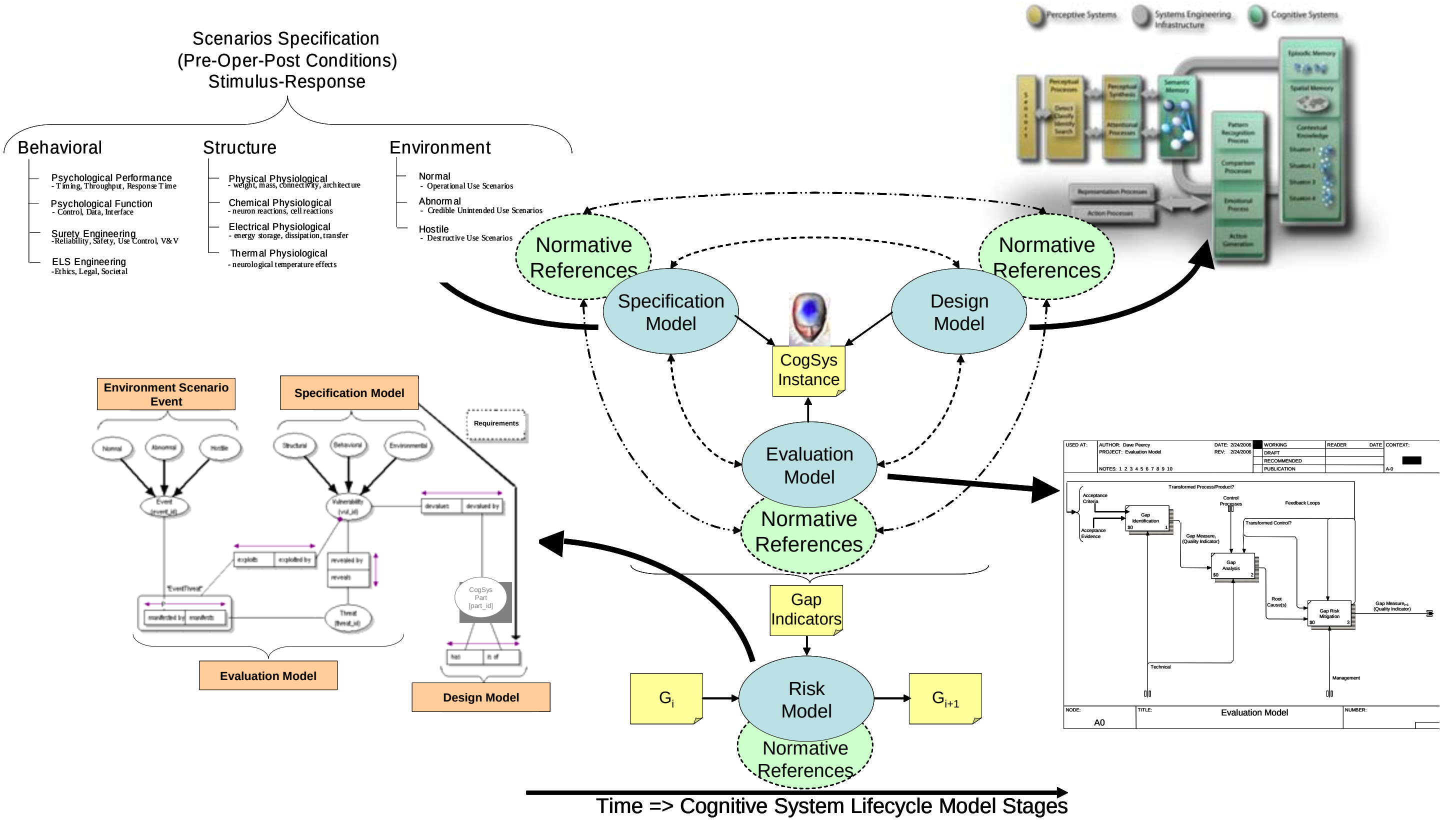


Figure EX-1. Surety Engineering Framework for Cognitive Systems

A process for applying the Framework to cognitive system implementations is specified with the following major steps:

Step 1: Classify Cognitive System Application

Step 2: Derive Cognitive System Implementation Models

Step 3: Determine Potential Areas of Risk

Step 4: Conduct Implementation Activities

Step 6: Conceptual Model Framework Feedback

A risk-informed decision process is defined that integrates a scenario-driven approach to determine potential design and/or implementation gaps derived from the systems/surety engineering activities. Typically, the scenarios are representative of operational environments and use, and are designed to cover both intended and unintended use. Identified gaps represent potential vulnerabilities that could result in technical, ethical, legal, and/or societal risks if threats can take advantage of the vulnerabilities during credible operational scenarios.

Several case studies are presented to illustrate application of the Surety Engineering Framework for Cognitive Systems. The case studies are not comprehensive, but provide enough depth to understand the potential value of this research.

One case study was applied to a Non-Invasive Cognitive State Detection Sandia LDRD project, funded in FY2005. The project explored the creation of a computer workstation capable of integrating a number of non-invasive sensors, as well as Sandia's cognitive modeling capabilities, for the purpose of extracting a user's current cognitive state. This project resulted in the ability to use existing models of users or create new user models to provide real-time system adaptation to the user. The ability to automatically capture a user's cognitive state allows for the development of cognitive systems that are adaptable to an individual user and allows for the evolution of more accurate models of the individual. The Surety Engineering Framework models are described in terms of existing information about this project, the maturity of the project results is outlined, and potential gaps and risks are identified.

Another case study was applied to a the Long-Range Iris Recognition in Nonideal Conditions Sandia LDRD project, funded for FY2009, that aims to create a high-accuracy, high-throughput iris recognition system that works with subjects and multi-meter distances. The project will explore new research and development in adaptive optics and software algorithms for iris recognition in non-ideal conditions. The integrated system will provide a high-accuracy, high-throughput, multi-meter distance iris recognition of both cooperative and uncooperative subjects. In this case, the Surety Engineering Framework models were applied to the conceptual aspects of the project to determine how well the project requirements and expected results were specified, how well the design concepts for the project appeared to reflect the fidelity of the specification requirements, and whether the proposed project evaluation methods addressed the conceptual aspects of the Framework's Evaluation Model. Applying the Framework in the early phases of a research project is expected to identify a more comprehensive set of potential gaps that can be addressed early and mitigated as necessary; more effectively communicate the risk level and maturity; and capture the issues/concerns that could be transitioned beyond research phase.

The first case study illustrates the application of the Framework to a research project that has been completed, while the second case study illustrated the application of the Framework at the Early Concept stage of a program's lifecycle. Since both cases are early research projects, each project will typically have a large quantity of gap indicators and the associated Cognitive System Maturity Matrix is expected to be at a low maturity index level. The goal would be to apply the Framework as a preventative mechanism to reduce the likelihood that identified gaps would be propagated into full-scale development and production products – as well as reduce gaps during the new t4echnology research and conceptual phases.

Several appendices of this report provide specific background information to support the application of the Framework and Process. These appendices address cognitive systems and surety engineering topics as summarized below.

I Existing Cognitive System Architectures And Emerging Technologies

This appendix describes existing cognitive system architectures as well as emerging cognitive neuroscience and related technologies. Nineteen different existing architectures are briefly summarized in terms of categorization as emergent, symbolic, or hybrid. These architectures represent various fidelity instances of the Framework Design Model. Some general conclusions are derived from the various architecture reference material as it relates to the Framework Design Model and associated normative references.

In addition, this appendix provides discussion points as to how the Framework addresses some of the key findings in an important 2008 pre-publication report by the National Research Council on Emerging Cognitive Neuroscience and Related Technologies. These key findings are directly addressed by the Surety Engineering Framework for Cognitive Systems – even when the systems of concern are not specifically cognitive. The bottom line from this report states:

> *"Cognitive neuroscience and its related technologies are advancing rapidly, but the IC has only a small number of intelligence analysts with the scientific competence needed to fully grasp the significance of the advances. Not only is the pace of progress swift and interest in research high around the world, but the advances are also spreading to new areas of research, including computational biology and distributed human–machine systems with potential for military and intelligence applications. Cognitive neuroscience and neurotechnology comprise a multifaceted discipline that is flourishing on many fronts. Important research is taking place in detection of deception, neuropsychopharmacology, functional neuroimaging, computational biology, and distributed human-machine systems, among other areas. Accompanying this research are the ethical and cultural implications and considerations that will continue to emerge and will require serious thought and actions. The IC also confronts massive amounts of pseudoscientific information and journalistic oversimplification related to cognitive neuroscience."*

Due to the complexity and extensiveness of the cognitive neuroscience and related technologies research, a systematic approach such as specified by the Surety Engineering Framework for Cognitive Systems is needed to:

(1) separate out pseudoscientific and over-simplified information (e.g., non-evidence-based research and research whose evidence does not support its claims),

(2) integrate ethical and cultural implications and considerations.

(3) identify the maturity of a multitude of emerging technologies from around the world,

(4) address serious military and national security challenges, and

(5) augment/improve the technical capabilities of the technology warning methodology.

To address these new technology concerns it is necessary to adopt a common framework that can be scaled to a large variety of applications, incorporates a systems engineering discipline, applies known as well as innovative methods and techniques for verification and validation of the technical requirements for reliability, safety, and security, and incorporates a disciplined/engineering approach to elicit and mitigate risks due to ELS concerns. The Surety Engineering Framework should be able to address these challenges as the approach is applied to specific applications and the Framework is improved and evolved.

II Cognitive Systems and ELS Risk Research

This appendix contains descriptions of research related to Cognitive Neuroscience Inspired Models in terms of normative reference specification model concepts, design model concepts, and mathematical modeling concepts.

In addition, this appendix describes cognitive system risks in the areas of ethical, legal, and societal principles. The discipline of surety engineering offers a rigorous and systematic approach to the identification and analysis of the spectrum of risks potentially triggered by the development and dissemination of cognitive systems and neuro-technologies. A substantial number of these risks are likely to be in the spheres of law and ethics. The surety methodology outlined in this report provides a framework in which technology developers can be prompted to anticipate legal and ethical concerns associated with their work and to do so beginning with the basic research stage of a project and continuing on with increasingly detailed analysis as a product or process is offered to commercial and government customers. Important legal and ethical questions cut across at least six general areas, and discussions and examples are illustrated in these areas: Responsible Science; Privacy; Informed Consent And Control; Public Dialogue; Human Enhancement; and Security. An example of cognitive system ELS and technical risk perceptions is presented in reference to the use of the functional Magnetic Resonance Imaging (fMRI) technology.

III Surety Methods and Technologies

This appendix contains some summary information regarding several surety methods and technologies that are normative references for the Surety Engineering Framework Evaluation Model. Topic areas covered include:

(1) Design For Reliability

(2) Design For Safety

(3) Design For Security

(4) Modeling And Simulation And Computational Analysis

(5) Quantification Of Margins And Uncertainty

(6) Experimental Design

(7) Verification And Validation

IV Educing Information Research

This appendix includes information on educing information concepts, including a description of detection deception applications and a description of general surety application strategies. This is a significant area of interest for National Security and Military applications.

One Sandia project on “Modeling Aspects of Human Memory and Reasoning” illustrates a cognitive system research project where the design model represents a higher fidelity version of the framework’s design model. Several project research areas support Educing

Information research. An analysis of the project’s specification, design, evaluation (particularly the V&V aspects), risk, and maturity models is provided from the viewpoint of the surety engineering framework.

Conclusions

One potential external customer of this Framework provided the following insight in regard to the Surety Engineering Framework for Cognitive Systems after an extensive presentation and discussion of this effort:

> *”This is an elegant model as you can change the nature or acuity of any element within the framework without changing the framework itself. The strength of this framework is the framework itself as it can be scaled for any size project or any data type.”*
>
> *“This framework provides the constraints required to address the slippery slopes of neuroethics, engineering the ethical, legal, and societal issues alongside other technical risks. This could be used a s certification process for all who work on neuroethics.”*

Hopefully the results of this research effort fulfill the insight provided in these statements.

Perhaps one of the most important characteristics of the Surety Engineering Framework for Cognitive Systems is that every implementation instance will provide lessons learned and updates to the Framework’s normative reference information and models. The framework itself can evolve as more information is obtained, normative references are improved, conceptual models are improved, and processes and tools for implementation are developed.

Recommendations

To determine how well the Framework and Process realize the expected benefits it is necessary to apply this research to actual cognitive system projects. Since such projects tend to be complex, it would be useful to apply the Framework and Process to a variety of projects in various stages of development and implementation. The Framework and Process are easily scalable to the life cycle stage as well as to the level of complexity and project size.

Specific Recommendations include:

(1) Apply the Framework and Process to one or more in-house Sandia and/or external customer cognitive system research efforts,

(2) Apply the Framework and Process to one or more military and/or national security projects addressing a cognitive system component development,

(3) Evolve the Maturity Matrix concept to include more definitive psychology and physiological cognitive system theoretical information, and

(4) Continue to research the concept of ELS engineering and determine what it means to apply the surety engineering approaches of QMU, V&V, and other such methods to this rather more subjective yet essential area.

1. INTRODUCTION

1.1. Purpose

Sandia has a responsibility to serve as the innovators who give first priority to risk identification, assessment, and mitigation strategies for cognitive system applications that improve military capabilities and national security. Cognitive science research investigates the advancement of human cognition and neuroscience capabilities. Addressing technical risks associated with these advancements can counter potential program failures, legal and ethical issues, constraints to scientific research, and product vulnerabilities. Survey results, focus group discussions, cognitive science experts, and surety researchers concur technical risks exist that could impact cognitive science research in areas such as medicine, privacy, human enhancement, law and policy, military research, and national security[1].

The purpose of this report is to provide the results from a short late start FY08 LDRD project titled "Investigating Frameworks for Application of Surety Methods to Reduce Development and Operational Risks of Cognitive Sciences and Technologies". The results include:

(1) a conceptual framework that can be applied to cognitive systems to incorporate surety methods in order to reduce the technical risk to cognitive system research and development and deployed operational applications;

(2) a process for investigating potential risks when applying the framework to specific use scenarios; and

(3) multiple case studies and research information to illustrate the use of the framework and process.

Although this research has some focus on military and national security areas, the framework and process can be applied to any cognitive systems and are scalable to component, subsystem, and system levels. The framework and process can also be applied at various stages of a cognitive system evolution – from early research through production of a commercial product. Surety methods such as safety analyses with failure modes and effects along with fault tree analysis, security cryptographic methodologies, and reliability/probabilistic methods offer a set of promising tools for analyzing and mitigating technical, ethical, legal and societal risks represented by cognitive systems.

Surety methods provide safeguards, improve verification and validation, and support technical risk mitigation. Sandia is a national partner in the research of and national policy for development of cognitive sciences and technologies in support of DOE's Office of Science Strategic Plan.

1.2. Scope

Researchers from Sandia and the University of New Mexico (UNM) collaborated on this project to characterize the cognitive systems framework, process, and illustrative case studies. For the purposes of this research, the definition of Cognitive System is:

[1] SAND2006-6895, "Investigating Surety Methodologies for Cognitive Systems," D.E. Peercy, W. L. Shaneyfelt, E. O. Caldera, T. P. Caudell, K. Mills, November 2006.

> *Cognitive systems are implementations of technologies that utilize as an essential component(s) one or more plausible models of human cognitive processes.*

The term “cognitive” is used in a broad sense throughout this report and is reflective of the cognitive sciences in general. This is similar to the approach taken in (NRC-2008[2]) to refer to “cognitive” as “psychological and physiological processes underlying human information processing, emotion, motivation, social influence, and development. Contributions from directly related cognate disciplines include behavioral and social science disciplines as well as contributing disciplines such as philosophy, mathematics, computer science, and linguistics. For our purposes, we believe it is also critical that the engineering area is added to this list of contributing disciplines, particularly surety engineering. Our concerns are also broadened into the areas of ethical, legal, and societal issues associated with such cognitive systems and related technologies – because they address who we are as human beings, which is primarily determined by our cognitive state.

Cognitive systems have one or more functions that model a human’s cognitive tasks. Such systems may be implemented as a computational system, a biological system, or some combination. Cognitive systems include implementations that advance/augment human cognition, simulate human cognitive tasks either for understanding or operational use, or perhaps limit/reduce human cognition capabilities. The “plausible” model means the cognitive system includes a realistic representation of the human cognitive process (at least some part) based on literature from psychology or neuroscience.

Some areas of potential applications for the surety engineering cognitive system framework and risk decision process include:

- Detection, recognition, analysis, and forecasting of human behavior and performance
    - fMRIs for deception detection
    - fMRIs for Neuromarketing
    - Cognitive state detectors within an office environment, vehicle, or aircraft
    - Cognitive models to represent an individual’s decision-making abilities and thought processes
- Machine representation and application of human knowledge and experience (synthetic subject matter expert)
    - Robots embodying cognition
    - Cognitive agents representing an individual
- System adaptation to the knowledge, skill, situation awareness, or intentions of individual operators or teams of operators
    - Dynamic cognitive models with the ability to learn

[2] National Research Council, “Emerging Cognitive Neuroscience and Related Technologies,” pre-publication copy of the Committee on Military and Intelligence Methodology for Emergent Neurophysiological and Cognitive/Neural Research in the Next Two Decades, National Academies Press, Washington DC, 2008.

- Preservation and transfer of knowledge and experience
    - Resources comprised of static cognitive models that represent subject matter experts
    - Training systems embodying cognitive models that represent subject matter experts
    - Neural implants to augment impaired or healthy cognitive states
- Aides to human attention, memory, situation awareness, decision-making, and other cognitive functions
    - Neural implants for communicating to devices
    - Neurotherapies to treat behavioral/learning conditions
    - Cognitive models used as personal assistants
    - Neuropharmaceuticals to increase focus and awareness
- Technologies in which human-machine interaction are vital to the performance, safety, and security of systems
    - Adaptive cognitive models that represent subject matter experts
    - Training systems embodying cognitive models that represent subject matter experts
- Training for jobs or tasks in which human interaction under unpredictable and stressful conditions is essential to success
    - Augmented cognition for advanced training

1.3. Motivation

One might ask – just why is there a need for such a framework as described in this report? How precisely will the results of this research – if fully developed – benefit cognitive science research and cognitive system technologies? Don't we already know how to do systems engineering and apply technical processes and practices that assure our products are safe, secure and reliable with no ethical, legal or social concerns?

The breakthroughs neuroscientists are experiencing and anticipated to achieve in the next several years are expected to have a significant impact on society, transforming the way we learn, heal, recall, communicate, and even think. Embedding cognition in machines is advancing capabilities that augment human performance in ways that will empower us to do tasks significantly more effectively, efficiently, and accurately. While beneficial to our society, these emerging cognitive technologies will have not only safety, security and reliability challenges, but also ethical, legal and societal (ELS) issues that must be addressed.

There are unique concerns surrounding cognitive science and its technologies. Often, these technologies are intended to interface directly to the brain to augment a person's cognition. How will such a technology affect a person's identity? How will we react to using these technologies for the purpose of enhancing healthy minds beyond "normal"? What will it mean to enhance ourselves only some of the time? Will employers, prison security guards, militaries, legal courts, and educators mandate the use of such enhancement? Addressing ethical, legal, and societal/sociological (ELS) concerns associated with cognitive systems that may define or alter who we are can be critical to the development of acceptable cognitive system technologies as well as provide evidence and mechanisms to counter unacceptable cognitive systems.

Autonomous cognitive entities embodied in machines or robots can be capable of making independent decisions and acting as free agents. These cognitive entities might be based on the cognitions of specific individuals. Who is legally and morally responsible for the decisions made? What validation and verification process will determine the reliability of their actions?

The threshold of this neuroscience revolution is here and the issues of concern are likely to overwhelm us before the legal courts can understand the ramifications, the regulations can be established, and society can grapple with the moral dilemmas. Addressing such issues as an afterthought is now a lesson learned by other sciences in recent history. Sandia has been discussing and addressing these issues for the past four years at workshops and conferences, as well as with policy makers and US Government agencies. Leveraging our expertise in cognitive systems and surety science, we are well positioned to inform, advise and lead by example.

A mechanism is needed to provide a responsible, proactive approach to developing cognitive system technologies. Just as with any technology of high consequence, we must ensure that they will be safe, secure and reliable with accidental or unintended uses identified and addressed.

The Surety Engineering Framework for Cognitive Systems described in this report is targeted to provide a foundation based on systems engineering, surety science and risk management to support prevention of issues we can foresee and preparation for those unseen.

In the area of human cognition, we have much to learn about the variabilities within the relatively “known” areas of neuroscience; and the “unknown” areas remain substantial. Our understanding of consciousness, emotion, creativity, and estimation of confidence are just a few of many areas yet to be uncovered. This vast arena of undiscovered territories of the mind leaves us with only a minimal grasp of both the intended and potentially unintended application uses of cognitive systems. Not knowing what to expect presents unique problems for technical assurance, and also has a direct impact on areas of risk that encompass ELS concerns. Hence, the prime benefit of the framework and associated risk-informed process is to provide quantitative and qualitative information on what we do know and what we do not know. This includes the technical and ELS risks that a given cognitive system and its underlying technologies might possess.

The framework and process are scalable to the project application, component/subsystem/ system level of focus, as well as the full product lifecycle investigation. This allows for addressing risks during research and development and preventing (or at least acknowledging) residual risks in operational products. The framework content at this early stage contains primarily examples of how the framework might be applied, but provides for its own evolution and improvement as normative standards, best practices, state-of-the-art tools and other applications of cognitive systems are discovered. Some specific aspects that provide motivation for use of the framework and the process because they address concerns indicated in the above paragraphs include:

(1) This framework provides a structured approach to examine and address the spectrum of risks associated with a project within any and all phases of the project's life cycle.

(2) This framework will serve as a tool to communicate the spectrum of identified risks and risk mitigation strategies.

(3) Using this framework will allow risks to be addressed concurrently with cognitive system research and development as opposed to reactively after problems arise.

(4) This flexible framework structure allows for the dynamic changes involved in any project. As a project changes directions and scales in size, this framework will change with it.

(5) This framework will create a document trail to describe what project risks were identified and how they were addressed throughout the project's life cycle.

(6) This framework will clearly communicate what risks are known, what risks are addressed, and what risks are not addressed. Even surmises about unknown risks can be documented to the extent they can be imagined.

(7) The strength of this framework is that it can be applied to a unique cognitive system project, as well as a general class of cognitive systems. In two parallel paths it pushes the limits on neuroscience while providing a social model that addresses social concerns.

(8) Normative references are the "truths" we rely on, including processes and standards, to develop technologies. While these "truths" are largely variable, the set of normative references will continually grow and change as research matures our understanding and knowledge. Thus, it is the structure of the framework where its strength lies, not in the content that will change over time. Currently, the availability of neuroscience normative references is minimal, but expected to evolve as we continue to learn more about the brain.

(9) The overview of project risk areas (see Figure 1-1 below) illustrates a clear, concise way to responsibly recognize and communicate the identified risks to the project team, management, and stakeholders. The project success quadrant is clearly the desirable outcome, but as much information as necessary about the other three quadrants is needed to ensure project success. This is the ultimate strength of the surety engineering framework for cognitive systems – identification of what we know, what we do not know, and how confident we are in that information.

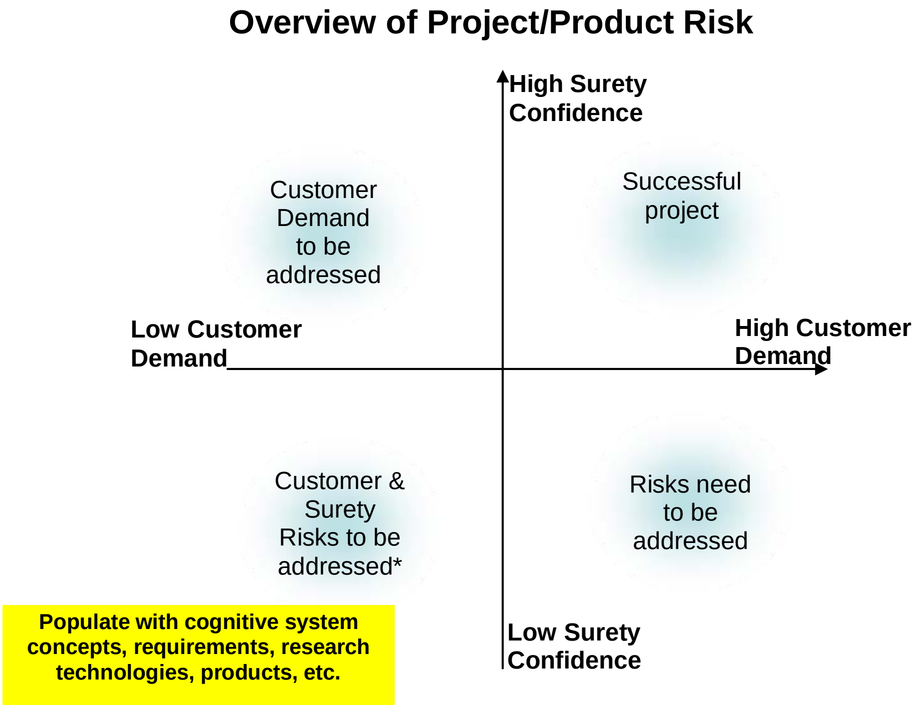

Overview of Project/Product Risk
High Surety Confidence
Customer Demand to be addressed
Successful project
Low Customer Demand
High Customer Demand
Customer & Surety Risks to be addressed*
Risks need to be addressed
Populate with cognitive system concepts, requirements, research technologies, products, etc.
Low Surety Confidence

Figure 1-1. Project Risk Quadrants Supported by the Framework

1.4. Audience

The audience for this report includes anyone who might have an interest in understanding the risks involved in the research and development of cognitive systems, how those risks can be comprehensively addressed, and potential safeguards to consider. This audience includes: management/policy makers; researchers/developers in the fields of cognitive science, neuroscience, surety, psychology and professionals addressing ELS issues; novices interested in learning about cognitive systems; and, ultimately customers/users of cognitive system technologies. The ultimate recipients of the benefits of this research are all of us – the public stakeholders that represent the reality of human cognition. Some potential benefits for this audience might be:

<u>Management/Policy Makers Audience</u> This framework provides an at-a-glance view of the risk areas that have been identified, the extent to which they have been addressed, and the real and potential gaps. This overview of risks can provide a focus for decisions such as where further funds should be applied, how information should be prioritized to the customer, and what public forums should transpire. The framework, when populated for a specific application, provides management access to related government regulations, import/export control laws, international trade rules, and so forth. Identification of required tradeoff studies and impact analyses will clearly be made evident and funding decisions can leverage the collected data.

<u>Developers and Training Audience</u> This framework provides a means to comprehensively specify and address the various risk dimensions related to cognitive systems. Developers are able to identify risk areas, analyze these areas, and explore mitigation strategies. They can easily identify and access applicable references and collaborate with others on discussion

topics of interest. Developers can select framework components as required by the rigor of their project. A cognitive system application developed for a customer will likely utilize more framework components than a research project with no clear application intent. As a training tool this framework can provide a comprehensive view of the risk areas related to cognitive systems, as well as resources available to access. It is intended to broaden an individual's understanding of the issues and considerations surrounding cognitive systems from a surety, ethical, legal, and public policy point of view.

Customers Audience This framework will provide customers reports on identified risks, analyses, and proposed mitigation strategies. The comprehensive view can provide assurance as to the issues that can be addressed and what future issues might be on the horizon as the technology matures or customer base expands. A maturity index will be developed for customers to comprehend the maturity of the cognitive system technology and ELS implications. The framework can be used to facilitate discussions on funding options and future growth paths. Customized reports will relay pertinent information in a concise and efficient format.

1.5. Research Technical Approach

The research technical approach for this effort is summarized in the following steps:

(1) investigate existing frameworks for cognitive systems as well as potential examples of existing and/or futuristic cognitive systems for which this research might apply (see Appendices A, C, D, E, F);

(2) derive a general framework for modeling the specification, design architecture, evaluation for verification and validation, and quality/risk indicators of cognitive systems (See Section 3);

(3) incorporate requirements within the specification model that address principles of ethics, legal, societal, and surety that address concerns identified in the cognitive system risk spectrum (See Section 3.2, Section 5 case studies, and Appendix F);

(4) incorporate existing cognitive system designs within the cognitive system architecture model (See Section 3.3, Section 5 case studies, and Appendix F);

(5) incorporate surety engineering methods such as safety, reliability, and security within the evaluation and risk models to address potential science-based vulnerabilities in the development and/or operational use of cognitive systems (See Section 3.4, Section 5 case studies, and Appendices E and F);

(6) apply a risk-informed decision process that can be applied to manage and hopefully reduce the risk of identified cognitive system vulnerabilities (See Sections 3.5, 3.6, 3.7; Section 4; and Section 5 case studies); and

(7) develop a case study that illustrates the value of and how to apply the derived general framework and risk-informed decision process (See Section 5 and Appendix F).

2. BACKGROUND

The focus of this research relies upon the collaboration of two areas of expertise within Sandia: Cognitive Systems and Surety Systems. While Cognitive Systems is a relatively new area of science for Sandia, systematic approaches to surety such as reliability, safety, security/use control, verification and validation, and human factors have been in place for multiple decades. Recognizing both known and unknown risks are involved while pursuing the cognitive systems research and developing technologies, an obvious prudent step is to mandate understanding the risks and then determining deliberate methods to mitigate them. Working in partnership with Sandia's surety experts, cognitive systems developers can attain a level of confidence in just how well the cognitive systems' technologies will operate as planned under both expected and unexpected circumstances. Addressing ethical, legal, and sociological (ELS) concerns associated with cognitive systems that may define or alter who we are can be critical to the development of acceptable cognitive system technologies as well as provide evidence and mechanisms to counter unacceptable cognitive system technologies.

2.1. Cognitive Systems, Science, and Technology

Sandia's Cognitive Science and Technology (CS&T) Program, established in 2006, creates a human-focused science and engineering base at the laboratory. The CS&T vision is to scientifically understand human brain, mind, and behavior to engineer technical solutions as applied to national security problems. This will enable the laboratory to provide answers to significant new challenges and threats as they relate to the human element of our nation's security. The human element is core to terrorism, rogue nations, dangerous weapon proliferation, and social unrest due to disruptive forces from changing societies, economies, and climate. The three CS&T scope objectives are: 1) scientific computing; 2) sensing and imaging; and 3) surety science and engineering.

The focus of Sandia's cognitive systems work today is on the development of computer models of human cognition that are applied to create unique technology solutions. By creating systems that embody cognitive characteristics of humans, we can take advantage of the basic strengths of humans and machines while mitigating the basic weaknesses of each. To date, Sandia's research has resulted in applications such as human performance augmentation, behavioral models and emulation, and cognitive state detection. There are many contributing disciplines such as: neuroscience, psychology, biology and physiology, social sciences, computing and mathematics, engineering sciences, physics, material sciences, and the relatively new field of micro and nano technologies.

The CS&T program pursues the development of cognitive systems technologies based on the belief that there are numerous positive impacts they could have on our national security. For example, a model describing how a human acquires knowledge through the process of reasoning, intuition, or perception can be customized to reflect an individual's knowledge and disposition toward various topics, tasks, technology, and people. The information gathered potentially might be used to detect deception in a potential terrorist, augment understanding of specific intelligence information, and create distributed human machine systems with potential for military and intelligence applications. However, concerns have been raised pertaining to such issues as the individual's privacy, legal ramifications for a model's intended and unintended use, and the technical aspects of verification and validation of the model.

Cognitive science researchers strive to attain the highest level of psychological and/or physiological representation of human cognition within realistic environmental conditions. Today, these representations are only partially achieved. Research in neuroscience, neuro-technologies, neuro-physiological processes, experimental psychology, human behavior, information processing, biometrics, and other related areas is providing knowledge, prototypes, analysis tools, and algorithms that can be incorporated into a cognitive system. Technologies are also employed that might be an input or response to a cognitive system. In this sense, such research and perhaps production products (e.g., functional Magnetic Resonance Imaging-fMRI) could become part of a cognitive system or part of the evaluation of a cognitive system.

It is important to understand that the research described in this report provides a systematic engineering framework that integrates assurance mechanisms to understand just how plausible the cognitive system research, development, and/or product realization is for its stated requirements. The resulting framework and application process only touches on the vast knowledge base that is cognitive system research. The framework and process application are illustrated primarily through examples. Just because the result is not highly plausible may be entirely appropriate and still an important contribution to cognitive systems. The questions to be answered are relatively simple to state, but somewhat complex to answer: what do we know and what don't we know?

The risks associated with the development of cognitive systems are related to the likelihood and impact of the occurrence of unwanted events associated with the use of cognitive systems. The question is whether surety technologies associated with such areas as safety, security, reliability and ELS can potentially reduce the risk (perceived or actual) so that the cognitive system research, development, and perhaps operational use might be considered acceptable for validated applications. Within a systems engineering approach, the surety engineering framework applies to the various cognitive system application research areas and the associated risk-informed decision process provides a view into potential technical, ethical, legal, and societal risks and the risk-mitigation maturity of those applications. Sandia has numerous existing cognitive system application efforts (see Figure 1-2) that can be formulated as instances within the context of the framework and even partial implementation of the risk-informed decision process. Examples of such instances are provided later in this report as case studies.

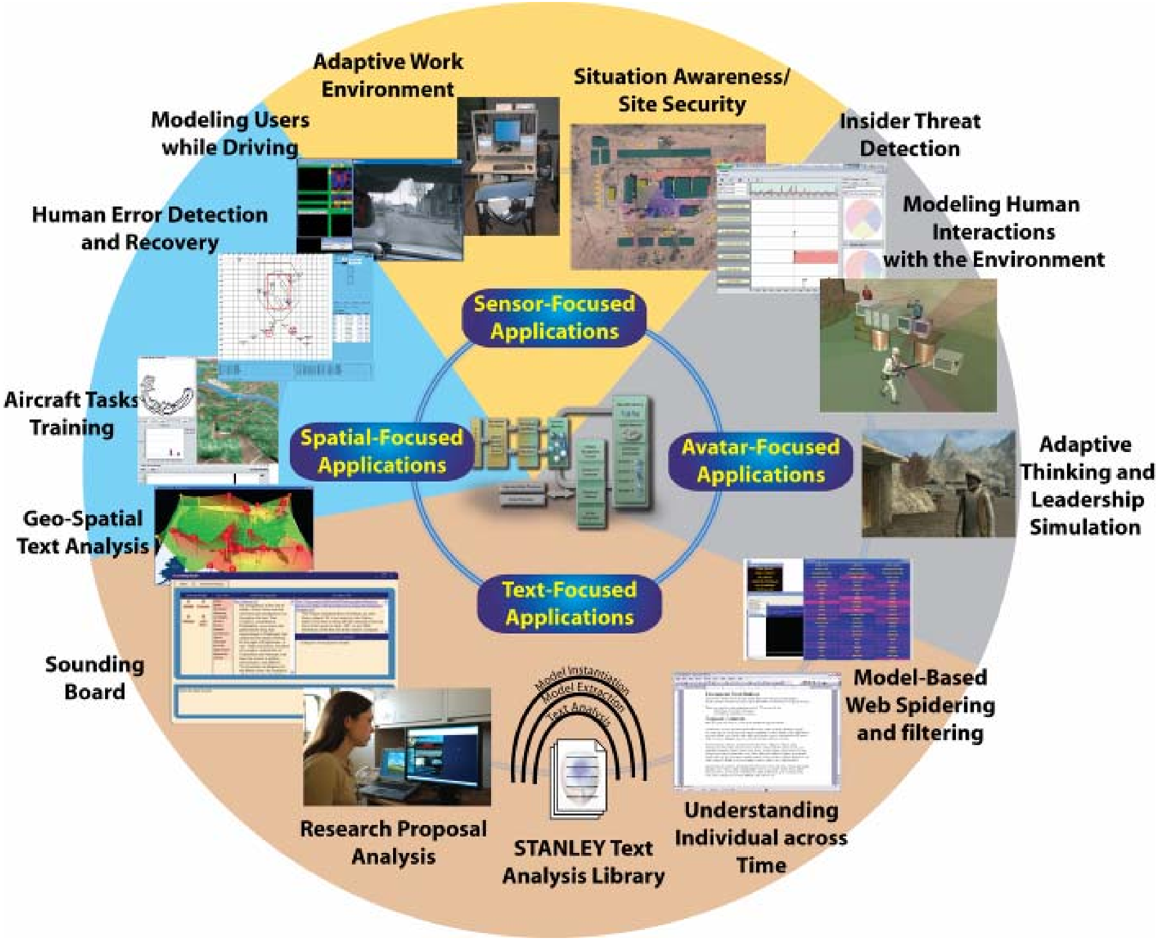

Adaptive Work Environment
Situation Awareness/ Site Security
Modeling Users while Driving
Insider Threat Detection
Human Error Detection and Recovery
Modeling Human Interactions with the Environment
Sensor-Focused Applications
Aircraft Tasks Training
Spatial-Focused Applications
Avatar-Focused Applications
Adaptive Thinking and Leadership Simulation
Geo-Spatial Text Analysis
Text-Focused Applications
Sounding Board
Model Instantiation
Model Extraction
Text Analysis
Model-Based Web Spidering and filtering
Research Proposal Analysis
STANLEY Text Analysis Library
Understanding Individual across Time

Figure 1-2. Sandia Cognitive System Application Research Areas

2.2. Cognitive System Risk Spectrum

The FY06 LDRD effort[3] derived a risk spectrum for cognitive systems along with a recommendation to develop the surety engineering framework. A summary of the surety risk areas as well as applicable surety methods is illustrated in Table 2-1.

Table 2-1. Cognitive Science Technology Risk Areas & Applicable Surety Methods

Risk Area	Rationale/Description/Concern	Surety Method(s)
Reliability (High Priority)	Human experience with technology is that 'all things break eventually'. Given the highly pervasive nature of potential applications, high levels of reliability will be necessary for them to be trusted. This will need to be demonstrated throughout their development, testing, and validation. In addition, the empirical nature of much of	Safety Principles Reliability: FMEA/FTA/PM/HF Methods/Sensitivity Analysis Risk Analysis: QMU Quality Methodology Ongoing monitoring efforts to detect adverse consequences early.

[3] SAND2006-6895, "Investigating Surety Methodologies for Cognitive Systems," D.E. Peercy, W. L. Shaneyfelt, E. O. Caldera, T. P. Caudell, K. Mills, November 2006.

	neuroscience, which currently lacks a broad theoretical basis, implies a high potential for unintended consequences.	
Privacy (High Priority)	Cognitive systems will incorporate significant amounts of individual information. Especially when used in the work environment, this raises concerns of access and use for purposes that may not benefit the individual. Further, this can extend to a sense of 'self-exposure', and an inability to control the degree of this exposure to others. Loss, theft, or unauthorized access bring consequently higher risks to the individual concerned.	Cryptographic Security can give capability to control access to the cognitive model. Control of the level of the cognitive model can also limit the 'personalization' of the model, and hence personal exposure through development and use of the model. Risk Analysis: QMU
Liability	Who is responsible in the case of malfunction? What constitutes informed consent in cognitive systems applications?	In tort law, responsibility is assessed according to the party's ability to mitigate the risk. This could be interpreted as the technology developer, the corporate entity, or the user, depending on circumstance. Due Diligence. Cryptographic Security Risk Analysis: QMU Quality Methodology Safety Principles Reliability: FMEA/FTA/PM/HF Methods/Sensitivity Analysis
Legal / Ownership / Intellectual Property	Questions include who owns a cognitive system, who controls its use, and who gains from it. Cognitive technologies extend the boundaries of possibility for humans, and also for machines. Courts may be called on to decide which individual rights apply in both of these cases. The technology, however, may become both ubiquitous and undetectable to the extent that enforcement of legal limits is not feasible.	Cryptographic Security Risk Analysis: QMU Quality Methodology Safety Principles Reliability: FMEA/FTA/PM/HF Methods/Sensitivity Analysis
National Security	To the extent that these technologies can be inexpensive, and require little infrastructure, they are highly attractive to 'bad actors'. Already in development, the US lead is not inevitable, and US policy decisions on appropriate use of these technologies will not necessarily have global sway. This quasi-obligatory technology development has the result that individuals perceive a sense of inevitability in the advent of the technology, which lessens their sense of having a true voice in its development.	Some issues can be addressed through security in development, and the design of system security features. The larger concern is one of international governance and policy. Safety Principles Reliability: FMEA/FTA/PM/HF Methods/Sensitivity Analysis Cryptographic Security Risk Analysis: QMU Quality Methodology
Hype and Backlash	Inflated claims, exaggerated fears, and genuine concerns over the implementation of cognitive systems in society may create a highly polarized spectrum of opinion that is prejudicial to balanced debate.	Surety methods may be able to provide convincing evidence that cognitive systems can be safe, reliable, and controllable. They may also contribute to the framing of a fact-based debate rather than a values-based debate. Public communication and discussion forums are non-technical methods to provide surety.
Dependency	Cognitive systems will exacerbate the increasing reliance of society upon technology, and may	Redundancy, system backups, and high reliability in systems will be crucial to

	contribute to an increasing separation of humankind from the natural world. Will this reliance cause human abilities to atrophy?	provide assurance of sustainability.
Diversity	Normalization results from one particular way of thinking becoming privileged because it is embedded in a widely used cognitive model. This also carries the risk that enhancement of one kind of cognition may come at the expense of other forms of cognition.	Risk Analysis: QMU Quality Methodology
Equity	Uneven access to cognitive technologies across socioeconomic groups raises the potential for a widening gap between rich and poor, both nationally and internationally.	These distributive justice questions are primarily addressed through public policy methodologies. Surety methodologies can help to achieve appropriate implementation in areas of the world with inadequate technical infrastructure.
Human Enhancement	There is a tension between the possibility for improved human performance, and the risk of irreversible and perhaps inappropriate changes to the course of human evolution.	Emerging technologies are creating unprecedented possibilities for shaping and changing the human future. This is an area of great uncertainty. Open discussions between scientists engaged in these technologies, members of the public, and other stakeholders will be vital for responsible development. Safety Principles Reliability: FMEA/FTA/PM/HF Methods/Sensitivity Analysis Risk Analysis: QMU Quality Methodology
Moral/Religious /Spiritual	Conflicts are increasingly emerging between faith-based beliefs and scientific discovery, fueled by opinions that such research is in conflict with faith-based values. Also, the relationship between the individual “self” and the cognitive model raises questions of identity, autonomy and human nature. Several participants expressed the sense that humans are irreducible; that there is a unique quality to human judgment and experience that cannot be replicated by technology.	Some faith-based concerns may be mitigated if such systems can be shown to be well delimited, and to have value for individual well-being. The maintenance of individual choice is important in this area. Attempts to integrate ‘ethical systems’ into cognitive systems face questions as to the particular ethical system to be selected. Nevertheless, an exercise of this type might offer a useful evaluation technique for systems under development.

A summary of the general conclusion from the survey group results of the FY06 LDRD concerning the likelihood of cognitive system applications posing ethical issues is summarized in Figure 2-1. It is clear that ethical issues related to the risk spectrum of Table 2-1 are perceived as likely.

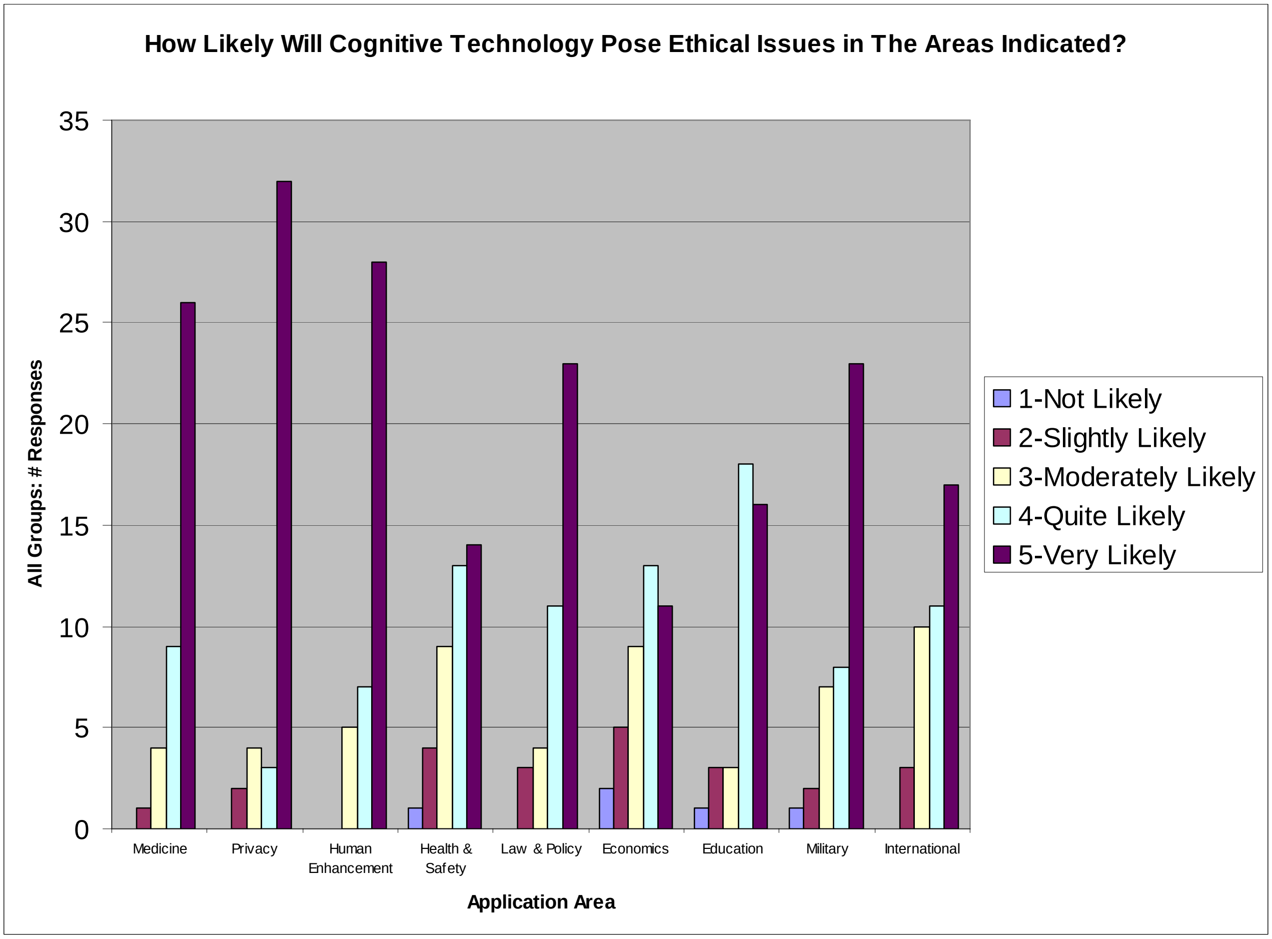


Figure 2-1. Summary of Perceived Application Area Ethical Risks (All Groups)

Because the risk spectrum addresses technical as well as ELS concerns, it is important that any cognitive system research, development, and/or product realization process address these concerns – the earlier the better. The following questions establish some of the criteria that the engineering of cognitive systems must address. The surety engineering framework and risk-informed decision process described in this report are expected to provide a mechanism for addressing these concerns.

Q1. Who sets the criteria? Who determines what the boundaries are and what mechanism or assessment is in-place to determine if a system is ethical?

a. The group setting the boundaries might have biases and not necessarily malicious biases, but latent educational, cultural, or social biases. What is good for one country is not necessarily good for another country.

b. Who will guard the guardians?

c. Are those selected to set the criteria from a homogenous element of the population with preconceptions and a resultant unintentional bias? Will this bias filter out an element of "diversity" that would make the pools of leaders less predictable, spontaneous, and representative of the society they're drawn from?

d. Would this selection process be fair? Is the intent of the selection process altruistic and universally accepted? If so, than the criteria might be fair and ethical. If you trust the group and you trust the model; then you trust the outcome to be fair.

Q2. How do we protect against misuse of a cognitive system application?

a. Misuse (meaning here a product used not for the original use intended) with good intent can sometimes lead to the innovative application of the product for other good intents. However, inverse is also true for bad intents.

Q3. How do we protect against abuse (use for a purpose that was not "good") of a cognitive system application?

a. What if a cognitive profile was compromised and fell into an adversary's possession? Could this give the adversary a clear advantage in that the leader's actions could either be anticipated or he could become easier to mislead?

Q4. How do we protect against accidental use of a cognitive system application?

a. What if cognitive profiles were accidentally released? This could be just as damaging as identity theft if obtained by unscrupulous parties.

Q5. How do we ensure the expected (and only expected) use of a cognitive system technology to cognitively enhance humans for the purpose of improving mental capabilities to process information, extract pertinent data, anticipate future events, and so forth?

a. Cognitive enhancement could be applied by a physical implant being emplaced in/on a person for the purpose of enhancing performance. Viewed as an "unnatural" way of enhancing a human's performance, it might incur more public scrutiny and might encounter more social and religious resistance.
b. What is to be done with enhanced individuals after they retire, become incapacitated, or are otherwise removed from their environment that required enhancement? Are their enhancements withdrawn?
c. Are there health risks involved due to invasive enhancement processes? Infection, natural cognitive deterioration due to dependency, mental, emotional, psychological?
d. Can the cognitively enhanced turn into an elite class with more than disproportional influence over areas they normally could not influence (e.g., economics, intimidation, etc.)
e. Will there be universal application available to all? If so, is it effective for all?
f. Is it ethical to not offer this technology if we have it?

Q6. How do we safeguard cognitive profiles (privacy)?

a. Is a "consent form", "cognitive content disclaimer", or some other administrative process required?
b. Should there be the concept of "cognitive liberty" or an unstated right to cognitive privacy to legally and ethically protect issues associated with the use of the material without the writer's consent.
c. Should there be a disposal plan or expiration date for cognitive models?

Q7. Can we use cognitive system technologies in a "fair" way for training that could improve educational opportunities, skill sets, technical abilities, and so forth?

a. Will everyone be given a fair chance in training to improve using these technologies?
b. Is it ethical to not offer this technology if we have it?

Q8. Is it ethical, legal, and socially acceptable to create cognitive system technologies for the purpose of selecting individuals via assessing and quantifying desirable traits?

a. Selection might not be a good use of the technology if the group selected would become too homogenous, predictable, and eventually "elitist"; not representative of the society the selection was drawn from.
b. Selection might preclude "out of the box" thinkers or other minority traits that add needed diversity and an element of unpredictability.
c. Selection might help us better place individuals in more appropriate positions.
d. Selection might be bias and produce unfair assessments.
e. Selection might stifle diversity.
f. Selection might foster too much conformity thereby producing a higher degree of predictability.

Q9. How do we get expert, as well as general public, acceptance of cognitive system technologies?

a. Ensure adequate safeguards are developed and put in place to protect individuals from abuse
b. Ensuring successful acceptance of the technology depends on the extent to which the technology is applied to life and death situations. Public might be more willing to take greater risks when human lives or national security are at stake. Acceptance of an emerging technology that wasn't applied to life and death situations (e.g., something that could make one perform a function better) might be a bit harder to employ as people might be less inclined to trade off what they have or what is known for possible benefits of the emerging technology.
c. The "intent" of the technology might make it either easier to accept or harder to deny. Favorable intent is doing the right thing for the right reason.
d. Applying lessons learned from introductions of other previous risky technologies might help mitigate similar problems.

Q10. How will cognitive systems change our definition of humanity?

a. Who gets the technology or benefits from the technology? Will particular individuals, cultures, militaries, races, or civilizations benefit? Will some be excluded from using the technology?
b. Will there be an asymmetrical development in some elements of the populace?
c. Will this cause a shift in power, an adjustment in social norms or stratification?
d. Will those who are cognitively enhanced be viewed as less human or just a more capable human?
e. Would more inexperienced people who are cognitively enhanced be accountable for higher expectations?

f. Would that which was once considered unethical become ethical?
g. Do we need to monitor the impacts on other individuals, groups, and societies?

Q11. How do we monitor the risks?

a. From an organizational perspective, attempting to apply the emerging technology should be monitored by a multi-disciplined cell that not only keeps abreast of the emerging technology, but also, any risks/ethical concerns associated with this in which all perspectives are welcomed and encouraged.
b. Industry wide requirement to create an ethical forum so that when the science is ready to go mainstream, there is a "self-regulating" entity in-place and therefore might preclude any governmental requirement to regulate the emerging technology to an extent so significant that technological progress is hindered.
c. Other forums – the press, legislature, and government regulatory activities that have a requirement to either keep the public informed or to safeguard the public.
d. Industry – marketing and publicity can help ensure that the public is informed and willing to accept the technology.
e. Watchdog organizations will be needed to validate and assess technology; inform; counter positions; and monitor progress.

2.3. Surety and ELS Engineering Principles and Methods

Surety engineering provides the processes, methods, and technologies that assure a product is reliable, safe, secure, and is able to be used as intended and not used in unintended ways. Surety engineering is part of the overall product systems engineering approach that provides adequate understanding of the margins and uncertainties that may limit the product application. As new technologies are integrated into product applications, it is essential to understand the potential use and misuse of those technologies and the resulting product. Since system/surety engineering encompasses the full product life cycle from concept to retirement – surety mechanisms must be considered as early as possible, even during research and development.

The surety areas of interest include: safety, reliability, and security but other cross-cutting areas such as human factors and on-going product support (where operational issues are identified and changes to systems are made) are other areas of interest. Sustainment of a usable system under an ever-changing technological landscape may be even more important than the original system development. A somewhat non-traditional "engineering" area is being proposed within this report to address the ethical, legal, and societal (ELS) concerns that are so important to cognitive systems – ELS engineering. One rarely thinks of ELS issues in an "engineering" sense, but the same engineering principles apply as for systems/surety concerns:

(1) What are the potential product application scenarios (intended and unintended variations) and associated ELS concerns?

(2) What are the ELS product requirements –privacy, legal, public/individual acceptance?

(3) How is a product design characterized to ensure ELS requirements have been addressed – authentication mechanisms, safety protection, denial under misuse, high consequence criticality analyses for reliability and fault tolerance?

(4) How is the product design evaluated to determine the margin and uncertainty in how well the ELS requirements have been addressed – quantification of margins and uncertainty, public acceptance index?

(5) What are the potential product vulnerabilities and threats for the identified product application scenarios that create gaps in satisfying ELS product requirements – verification and validation within a risk-informed decision process, capability maturity?

(6) What are the risks that ELS threats might take advantage of existing vulnerabilities – lack of protection mechanisms against unintended use?

The system/surety engineering areas have been significantly studied and applied within SNL weapon/weapon-related applications as well as for other technologies. This section briefly describes some of the normative principles in these areas and typical methods that can be used to assure cognitive system models and their implementation. Some thoughts on ELS engineering principles and methods are also summarized. Further details on surety methods, cognitive system technologies, and ELS risks are presented in the Appendices. Application of such principles and methods will result in the reduction of risk associated with cognitive systems. However, the challenges are significant because of the uniqueness of human cognition. Some of the unique characteristics of human cognition that require particular care in implementing cognitive systems include:

(1) The large body of unknown information about the human cognition psychological and physiological models;

(2) Identity of self - altering cognition differs from effects of physical or chemical changes;

(3) Attaining or superseding cognitive "normal" – altering cognition can be done for new purposes;

(4) Privacy of our minds – directly connecting to the mind to explicitly learn about someone;

(5) Trespassing in our brains – directly connecting to the mind to implicitly learn about someone; and,

(6) New "snake oil" for our brains – new dangers associated with "get smart, get focused, learn more" products and techniques.

2.3.1. Safety

Sandia has developed a strong infrastructure and process definition[4] that ensures systems are safe. Cognitive systems must also exhibit a strong verification and validation that they are safe. The key to Sandia's approach to safety is its attention to first principles. Cognitive systems may not be dependent on physics (or biological, chemical) principles, but clearly are

[4] DG10100/B, "The Process for Achieving Nuclear Weapon Safety at Sandia National Laboratories," Design Guide, Issue B, 2003.

dependent on the behavior of such systems. For safety purposes, cognitive systems should attempt to be developed using the following three principles of isolation, inoperability, and incompatibility as well as the implementing principle of independence.

o Isolation: critical components are separated from each other in a manner to preclude undefined interactions. Components that control safety-critical functions are isolated from other components.

o Inoperability: abnormal conditions cause the component to become inoperable in a safe, predictable manner, and before any isolation features are compromised. In hardware, inoperability also implies that the component does not become operable without a deliberate external reset. As applied to software design, these criteria can be implemented through comprehensive exception handling and fail-safe designs in critical components.

o Incompatibility: the interfaces among components are designed such that unintended connection cannot be made. Also, as with Isolation and Independence, the use of well-encapsulated components with a well-defined external interface definition may be applicable.

o Independence: stimuli for actions originate from and are handled by separate components. One implementation may be by redundant components with different designs that support a safety related task. As applied at a systems level, it implies an implementation that requires more than one failure of independent components before resulting in a safety hazard.

2.3.2. Reliability

Sandia has a strong reliance on actual experimental data to determine reliability measures. Specific methods such as Failure Modes and Effects Analysis (FMEA) and Fault Tree Analysis (FTA) support the specific analyses of system, subsystem, and component reliability. Probabilistic Methods (PM) is a promising method for determining reliability under conditions of uncertainty. The use of Quantification of Margins and Uncertainty (QMU)[5] as part of a risk-based approach to verification and validation decisions is a promising approach to understanding the fidelity of computational models such as part of a cognitive system. A Predictive Capability Maturity Model (PCMM) is being developed as a way to quantify how well computational models can predict accurate results. Such a model would be invaluable as a verification/validation approach for cognitive systems.

Reliability design principles and techniques include:

- Failure Mode Identification
- Lessons Learned
- Evaluation of Design Changes
- Reliability Improvement Analyses
- Design Concept Comparisons
- Iterative Optimization Analysis
- Assurance of Testability

[5] M. Pilch, T. Trucano, J. Helton, “Ideas Underlying Quantification of Margins and Uncertainties (QMU): A White Paper, ” SAND2006-5001, September 2006.

- Sensitivity Analysis
- Risk-based Decision Analysis

Most cognitive systems will involve the use of commercial components as well as development of custom components – for both hardware and software. Sandia has applied methods in the study of the reliability aspects of complex systems of custom and commercial products that are applicable to any systems, including cognitive systems.

2.3.3. Security/Use Control

Sandia has developed key methods and techniques to ensure their critical systems have adequate assurance of authorized use and protection from unauthorized access/use. State of the art cryptographic encryption methods have been developed and deployed within the requirements of the National Security Agency. Such methods and techniques clearly have application to cognitive systems where concerns such as privacy, ownership, and operational control are important. Some of the key elements of security include:

- Unauthorized access detection
- Authorized access initiation and verification
- Cryptographic system lock/unlock verification
- Disablement of system upon unauthorized access
- System reset on authorized access command
- Activity monitoring reporting

2.3.4. Human Factors

Cognitive systems, particularly the targeted ones for this short study, will have many human factors concerns. Human Factors (HF) engineering is the process of designing for human use. The objectives of this discipline are to reduce the opportunity for human error and to enhance the productivity of human-machine systems. Sandia does this by systematically applying information about human characteristics and behavior to the equipment, procedures, and environments in which people work. These same principles and skills can be applied to the development and use of cognitive systems. Some of the skills Sandia can apply include:

- Task analysis
- Human-computer interaction design and evaluation
- Equipment layout and facility design
- Evaluation of human performance in various settings
- Human reliability analysis
- Survey construction
- Anthropometry and physical human-system interface design
- Design of experiments and statistical data analysis
- Test and evaluation of human-machine systems
- Vulnerability analysis of safeguards and security systems

2.3.5. Surveillance – System Sustainment

Sustainment of a system in the context of support changes is a challenge that is addressed by Sandia's Surveillance program. A well-defined process is required to ensure a system maintains its operational capability, investigate system failures/faults, conduct root-cause analyses, and integrate upgrades/modifications into the operational products for complex systems. For cognitive systems to be effective, it is essential that a support concept is put in place and the inevitable stream of upgrades is effectively handled.

2.3.6. Ethical, Legal, and Societal Factors

Science-based ethics, legal, and societal modeling and engineering implementation of cognitive systems is similar to any discipline. Application of such principles and methods will result in the reduction of risk associated with cognitive systems.

Principles for ELS engineering have been defined[6]. Cognitive system developers will:

(1) apply their respective **established professional guidelines** as appropriate;

(2) proactively consider the **intended uses and impacts** of their specific technologies, as well as the potential for accidental use, misuse, and abuse;

(3) provide inherent **safety features** to the extent they consider reasonably possible to maximize the prevention of accidents, misuse, and abuse;

(4) proactively initiate **ethical discussions** among themselves and with the wider public;

(5) provide **human test subjects** with a clear understanding of the personal information acquired, stored, analyzed, etc. and how that information will be used;

(6) handle responsibly any **personal information** obtained from test subjects; and

(7) respect the **limitations of a cognitive model** as a representation of a test subject.

These Principles span privacy, safety, human test subjects, and application use. These principles also embody methods and techniques that can be applied during the product life cycle to ensure the degree to which the principles are met. Typical activities for a cognitive system or associated technology would be to conduct technical, peer review, and public review of product engineering information such as:

(1) purpose (emphasizing not just that it can be created, but that it fulfills a real requirement);

(2) capabilities;

(3) intended use and potential unintended use;

(4) operational environment;

(5) limitations (including what it will not do);

(6) known risks associated with use and deployment of the technology;

(7) unknown risks that can be speculated (aka the "known unknowns");

[6] W. Shaneyfelt, "Ethical Principles and Guidelines for the Development of Cognitive Systems," SAND2006-0608, Sandia National Laboratories Report, May 2006.

(8) performance metrics; and

(9) user scenarios.

Value scenarios[7] are an interesting extension of scenario-based design[8] that might be useful in evolving the concept of the Surety and ELS engineering methods into the human cognition applications. Value scenarios support critical, systemic, long-term thinking in current design practice, technology development and deployment. The key elements of value scenarios are: stakeholders, pervasiveness, time, systemic effects, and value implications. Scenarios are integrated within much of the surety engineering information throughout this report and are intended to include such concepts as "Value Scenarios".

[7] L. Nathan, P. Klasnja, B. Friedman, "Value Scenarios: A Technique for Envisioning Systemic Effects of New Technologies," ACM Conference on Human Factors in Computing Systems, Chicago, 2007.

[8] M. Rosson, J. Carroll, "Scenario-Based Design," in J. Jacko & A. Sears (Eds.), The Human-Computer Interaction Handbook:L Fundamentals, Evolving Technologies, and Emerging Applications," Mahwah, NJ pp 1032-1050, 2003.

3. CONCEPTUAL SURETY ENGINEERING FRAMEWORK FOR COGNITIVE SYSTEMS

3.1. Overview of the Framework

Quality is the result of managing vulnerabilities to a targeted risk. Whatever vulnerabilities exist in a system that can be exploited by threats will define the system risk as well as the resulting quality. When systems have few vulnerabilities that can be exploited by threats, the system will have a low targeted risk and high quality. In the case of cognitive systems, it is important to reduce the potential risks in the risk spectrum by both eliminating vulnerabilities and limiting the potential exploitation by a threat. Elements of quality assurance/systems engineering and quality assessment have been a major part of the Sandia culture. The integration of quality engineering principles within the system development process and the conduct of independent assessments to understand how well desired quality is being achieved are essential to achieve requisite system quality.

Cognitive systems, by the very nature of their applications, must achieve a reasonable level of quality. A generic systems/surety engineering framework architecture has been developed that would also apply to any system – but in this report is being applied to cognitive systems. Four models are part of the architecture: Specification Model, Design Model, Evaluation Model, and Risk Model, as illustrated in the Figure 2.3-1. In addition, a Maturity Model is applied appropriate to the life cycle activities to represent the plausibility of the cognitive system implementation with regard to identified risks as part of the risk-informed decision process.

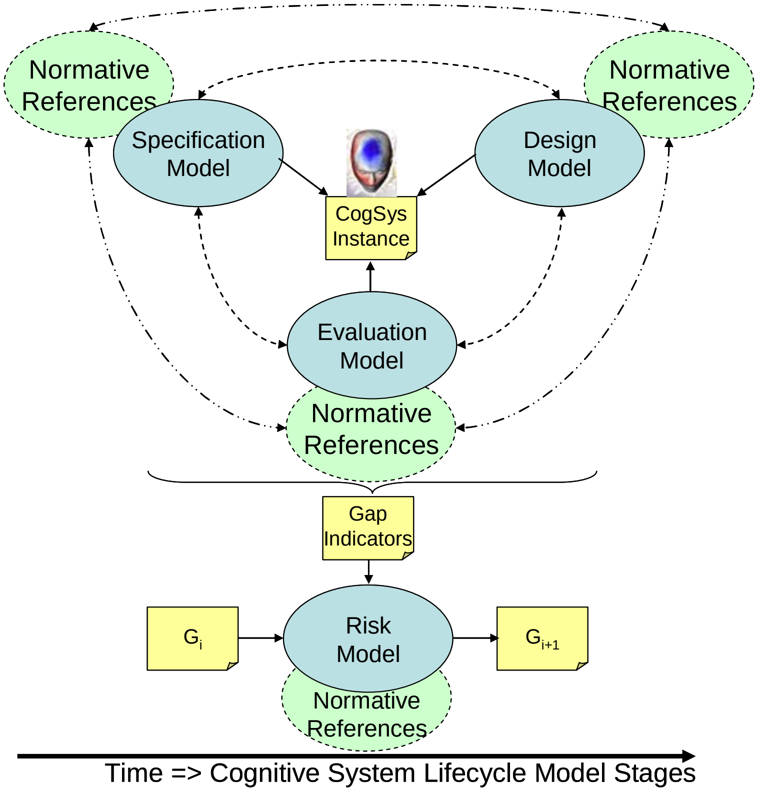


Figure 3-1. Cognitive System Conceptual Framework

A summary of the key terminology for this framework includes:

- **Environment Scenario**

  Sequence of activities through which the product (system, subsystem, component) is intended to satisfy its specifications in accordance with how it has been designed.

  Instance: intended and/or unintended scenarios for cognitive system use.

- **Normative References**

  Standards, historical evidence/lessons learned, and/or expert opinion that represents process and/or product best practices for any of the other elements of the framework.

  Instance: safety first principles, ethical principles, psychological theories and standard models, physiological structures and standard models, state-of-the-art tools (e.g., fMRI) that might be used as part of or for evaluation of a cognitive system.

- **Specification Model**

  Generic requirements (behavioral, structural, environmental) that address the class of products/processes (in our case, cognitive systems) within the scope of the quality framework.

  Instance: requirements of a specified cognitive system product/process.

- **Design Model**

  Generic architecture (physical/functional) that describes the class of products (in our case, cognitive systems) within the scope of the framework.

  Instance: design and processes used for a specified cognitive system product; use of tools (e.g., episodic memory model) as part of design.

- **Evaluation Model**

  Generic processes and methods that might be used to obtain measures of how well the requirements of the specification model are met by the architecture of the design model.

  Instance: specific verification and validation experiments with QMU analysis to obtain measures of how well the requirements of a cognitive system product/process are met; use of tools (e.g., Design of Experiment, Statistical Process Control, vibration/shock/temperature testing processes and equipment, fMRI) as part of evaluation methods.

- **Risk Model**

  Generic gap analysis processes and methods that might be used in a time/phase-dependent approach to determine the implications of the gap measures obtained by the Evaluation Model. Risk-informed approach to managing vulnerabilities to a targeted risk.

  Instance: risk-based analysis (potential threat, impact of threat/vulnerability occurrence, and likelihood of occurrence) of a cognitive system's lack of a safety theme implementation, or lack of an authentication mechanism to prevent exploitation of cognitive system privacy information, or a process gap in the conduct of external peer review of the ELS concerns for a new neuroscience technology; representation of the gap/risk indicators in a risk-informed prioritization based on the risk aversion threshold and perhaps the life cycle stage of the cognitive system.

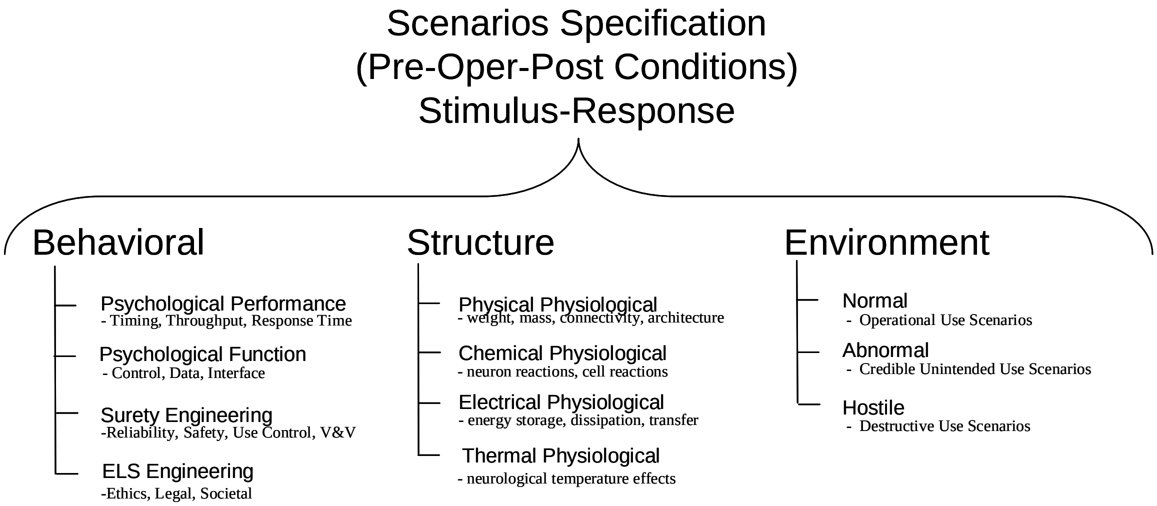
- **Maturity Model**

  Plausibility characteristics of the cognitive system (psychological, physiological, environmental conditions) as they relates to potential application use.

  Instance: research, early prototype development, full scale development and production, high consequence (regulatory) qualification application represent stages for which maturity/plausibility criteria might apply; measure of how well an episodic memory model actually represents human cognition.

3.2. Specification Model Details

The Specification Model for Cognitive Systems includes the specifications for the engineering processes and product requirements. These requirements are illustrated in the Figure 3-2.



Scenarios Specification
(Pre-Oper-Post Conditions)
Stimulus-Response
Behavioral
Psychological Performance
- Timing, Throughput, Response Time
Psychological Function
- Control, Data, Interface
Surety Engineering
-Reliability, Safety, Use Control, V&V
ELS Engineering
-Ethics, Legal, Societal
Structure
Physical Physiological
- weight, mass, connectivity, architecture
Chemical Physiological
- neuron reactions, cell reactions
Electrical Physiological
- energy storage, dissipation, transfer
Thermal Physiological
- neurological temperature effects
Environment
Normal
- Operational Use Scenarios
Abnormal
- Credible Unintended Use Scenarios
Hostile
- Destructive Use Scenarios

Figure 3-2. Cognitive System Conceptual Specification Model

The key components of the Specification Model are the Behavioral requirements, the Structural requirements, and the Environmental requirements – as well as combinatorial effects across these three components. These requirements are in the context of a cognitive system Concept of Operation (CONOP). This CONOP defines at a high level how requirements are to be satisfied within a specified set of scenarios and, within each scenario, the operational characteristics of the cognitive system prior to, during, and after the application of the scenario. The scenario might be for intended use environmental conditions (Normal) as well as unintended use environmental conditions (Abnormal or Hostile). The

Behavioral and Structural requirements may vary depending on the scenario and environmental conditions.

For the Behavioral component, the cognitive system requirements address the psychological performance and functional conditions that are to be satisfied by the cognitive system. In addition, system/surety engineering and ELS requirements are specified to ensure the cognitive system implementation will address those areas. At the conceptual level, such specifications might be represented by statements such as specified below, with more detailed explanation of what the statement means in the context of the normative references for human cognition functions.

Behavioral Model Basis

- *The model shall execute cognitive functions consistent with plausible psychological models of how humans think (Psychological Basis)*

For the Structural component, the cognitive system requirements address the physiological aspects of neuroscience technology representations of the brain physical, electrical, chemical, and thermal structure (macro and/or micro level) as well as the process representation of the brain inter and intra connectivity dependencies.

Structural Model Basis

- *The model shall execute cognitive functions consistent with plausible physical functioning of the human brain (Physiological Basis)*

For the Environmental component, the environmental/use scenario defines both the normal intended use as well as the potential abnormal unintended use and even hostile unintended use. These environmental scenarios are applied to specific Behavioral and/or Structural requirements that in combination must be satisfied. The criteria for acceptance may require certain evidence prior to the scenario, during the scenario, and/or after the scenario. As an example, there may be a cognitive system being used for educing information from a subject. The environment may require specified thermal/physical conditions and responses prior to the application of the educing scenario, specified thermal/physical conditions and expected subject responses during the educing scenario, and specified thermal/physical conditions and expected subject responses after the educing scenario. The cognitive system is required to work as specified during all three conditions. Validation of that requirement would be part of the evaluation model activities.

Environmental Model Basis

- *The model shall execute cognitive functions consistent with plausible environments for the behavioral and structural models (Environmental Scenario Basis in normal, abnormal, hostile environments)*

A normative reference for requirements data includes the first five data items/attributes below, but should be linked with the other four data items in what is termed a “requirements dimension”. This “requirements dimension” is important for identification of gaps during the evaluation process.

1. ***Requirement ID:*** unique identifier for tracing requirement.

2. ***Requirement Dependencies:*** source of requirement and link to any other requirements which are either dependent upon this requirement or upon which this requirement depends.
3. ***Requirement Statement:*** what is needed.
4. ***Rationale:*** clarification and interpretation of the requirement as appropriate; this is typically needed when translating rather vague customer requirements into some requirement statement or multiple statements that are unambiguous, testable, and so forth.
5. ***Measurement/Acceptance Criteria:*** what criteria determines if the requirement has been met – that is how do we know?
6. ***Verification/Validation Approach:*** scenarios, methods, techniques, tests that will be conducted to determine if one or more of the measurement criteria within the quality model attributes have been met (may vary over life cycle).
7. ***Actual Evidence:*** reference to definitions, models and documents from verification/validation results (may vary over life cycle).
8. ***Gap Indicator:*** difference between the measurement criteria and the verification activity results, as appropriate, with variance, confidence, and uncertainty margins.
9. ***Risk Inference:*** risk measure based on the Gap Indicator margin as judged by a Risk Agent. Such attributes can be applied to both product and process requirements.

The "Requirement Specification" may be included in a project plan (primarily for research), specification paragraph, state diagram, table, or other representation forms. It is usually necessary to develop a mechanism to capture the "requirement data" – for traceability and understanding of change impact. That "mechanism" may be an internally defined database, simple excel spreadsheet matrix, requirement management tool, or simply a requirements document. All the information can be directly included or referenced by links/pointers to the location of where such information is described (perhaps in more detail). A "Requirement Specification" may be the result of research, concept exploration, prototype development or a formal requirement elicitation, representation, and documentation process.

Some of the requirement statements[9] that may be generally descriptive of a personal cognitive system include:

(1) shall know what you know, including the underlying structure of your knowledge, and what you don't know;

(2) shall know what you do and how you do it, including the knowledge implicit in your actions;

(3) shall know about your past experiences and can properly place events within the context of past experiences;

(4) shall be able to apply your unique knowledge and experiences to interpret events in a manner consistent with how you would interpret the same events;

[9] see [FORSY-2005]

(5) shall be able to recognize when you have learned and how learning has reshaped your knowledge of the world; and

(6) shall know the consequences of your past experiences and the resulting sensitivities, and can anticipate how you will react to future situations.

3.3. Design Model Details

The Design Model for Cognitive Systems includes the conceptual architecture to implement the engineering processes and product requirements. The design architectural elements are illustrated in the Figure 3-3.

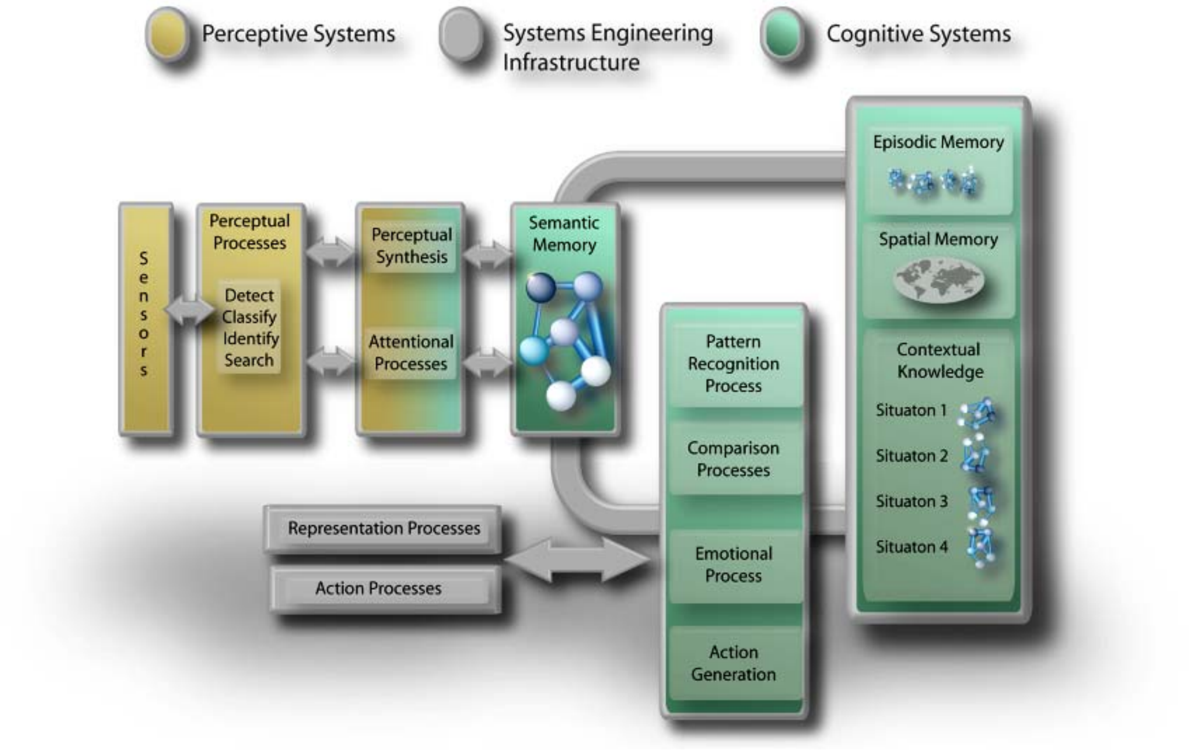


Figure 3-3. Cognitive System Conceptual Design Model

This specific Design Model is the existing Sandia Cognitive System model[10]. Although it is representative of what our framework requires as a conceptual design model, it is also recognized that there are many other existing frameworks (see Appendix C) that may be useful. One of the key aspects of our framework is that there is the expectation that it will evolve in its specific content, but have stability in the model concepts within the framework. For example, an alternative Design Model could be substituted for ours, but the overall framework of models would remain the same.

[10] C. Forsythe, M. Bernard, T. Goldsmith, Cognitive Systems: Human Cognitive Models in Systems Design, Lawrence Erlbaum Associates, Inc., 2006.

Each of the Design Model elements will be briefly summarized in the following paragraphs. A higher fidelity instance of this Design Model is illustrated in Appendix F as another example of the application use of the framework. The primary elements of the Design Model are illustrated in Figure 3-3 as:

(1) Perceptual Systems

(2) System Engineering Infrastructure

(3) Cognitive Systems

This existing terminology is not quite consistent with the overall surety engineering framework, since the complete Design Model is considered to be a cognitive system in our terminology. For purposes of this report – and as a simplistic view, the “cognitive system” within the Design Model represents the functions more commonly associated with specific cognitive functions. The “system engineering infrastructure” represents the functions more commonly associated with the physiology infrastructure that connects inputs, cognitive functions, and outputs of the cognitive system. And the “perceptual systems” represents the functions more commonly associated with the inputs to and outputs from the cognitive system functions.

3.3.1. Perceptual Systems

The perceptual systems include the sensors and perceptual processes that provide ways to gather and store information for cognitive function processing and make available/report results of that processing. The sensors are the typical visual, auditory, touch, smell, taste as well as non-typical mechanisms such as balance and temperature. The perceptual processes include perceptual memory mechanisms that allow for storing information related to volatile proto-objects, context to guide focused attention including abstract meaning and spatial arrangement, and the overall formation of coherent object recognition.

Sensory systems acquire information that is processed by perceptual systems to detect, classify, identify, and search for other existing relationships. The attentional processes provide object recognition and categorization of the results of the perceptual system associations and the perceptual synthesis of this information involves interaction with such processes as reasoning, emergent goals, context knowledge, and other such cognitive functions using the brain’s internal “systems engineering infrastructure” of neuron network and cerebral cortex functions that play a key role in memory, attention, perceptual awareness, thought, language, and consciousness.

3.3.2. Systems Engineering Infrastructure

The systems engineering infrastructures consists of the specific representation processes and the action processes that model the complex perceptual and motor processes and neuron communications network associated with the cognitive functions of the brain. At a macro level these models may only function to represent the “black box” interfaces with more plausible cognitive function models. At a micro level these models may represent the cellular reactions and electromagnetic interactions of a brain’s neural network.

3.3.3. Cognitive Systems Functions

The cognitive functions include the following elements[11]:

(1) semantic and contextual knowledge:

- semantic memory: abstraction of common features among related episodes and other obtained knowledge; knowledge of concepts; storage of knowledge about the "relatedness" of concepts; mathematically this is represented by sets with relations.
- episodic memory: integration of item familiarity and temporal-spatial context associated with specific experienced episodes; actual record of events stored; supports at least two cognition functions – contextual association of new episodes with existing episodes, and medium for learning mechanisms through recognition of recurrent events.
- contextual knowledge: recognition of concepts within a specific context for their understanding; different concepts may be recognized within the same context and the same concept may have multiple contexts such as time, place, objects involved, causal states, intended resolution

(2) spatial memory: spatial information in an imaginary or real environment; associations matched to stored information about the location, name, and function of other object(s) in the visual scene in terms of partial hierarchies; mappings and processing of spatial relation associations such as above, below, under, left of, right of, distance, direction, orientation, and motion concepts.

(3) action generation: monitor learning and performance; emotion assessment associated with episodic memories; perceptual-motor processes that result from cognitive function processing

- ▪ pattern recognition process: the process of associating patterns in terms of their context
- ▪ comparison processes: processes that monitor semantic memory and the concepts that are activated; triggered (more noticeably) when one or more concepts are activated in semantic memory that area inconsistent with current context(s); in addition, processes that may monitor other components such as perceptual-motor processes;
- ▪ emotional process: processing of emotions such as pleasure, anger, fear, anxiety, and disgust as well as their relationships with semantic memory and contexts within contextual knowledge.

[11] Bernard, Michael et al. "Memory & Reasoning LDRD Design & Testing Report," SAND2008-xxxx, 2008.

3.4. Evaluation Model Details

The Evaluation Model requires a process flow representation of how identified gaps are analyzed and decisions made concerning those gaps. To some extent this is a combination of typical Risk Analysis and Corrective Action processes. This process flow representation is illustrated in Figure 3-4.

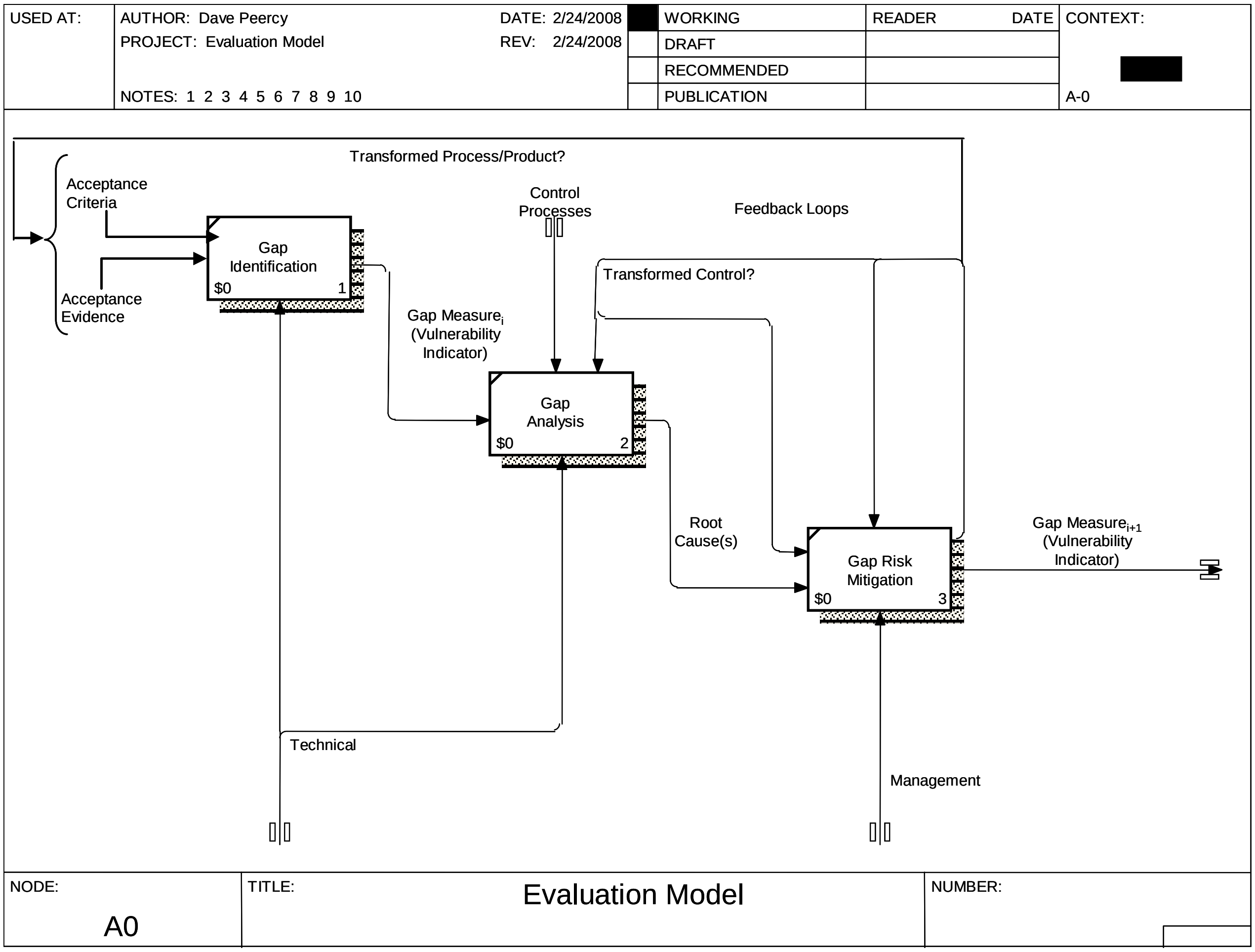


Figure 3-4. Cognitive System Conceptual Evaluation Model

The Evaluation Model provides visibility into specific risk indicators, risk mitigation results, and trends over the product life cycle. In addition, the Evaluation Model provides the opportunity for timely promotion of associated project risks to an appropriate management level where the risks can be most effectively resolved. The concept of the Evaluation Model is to provide enough information at any life cycle point so that risks can be identified, communicated, hopefully mitigated, and by design prevented from propagating to any production product.

3.4.1. Gap Identification

Risk indicators (gaps) are identified when the verification/validation evidence of a specification does not satisfy the specification's acceptance criteria. These gaps may occur in process or product specifications, and may occur at any point in the product lifecycle. Such gaps may be identified through normal verification/validation activities, or may arise through subsystem integration verification/validation activities, and/or internal/external assessments. Gap identification information should include:

- Specification ID:
- Product ID:
- Lifecycle Stage:
- Description of the Verification/Validation Activity:
- Description of the Verification/Validation Results:
- Risk Indicator Gap Measure:
- Gap Analysis Owner(s):

Once Gap Identification has been completed, this information is passed on to the Gap Analysis Owner(s).

3.4.2. Gap Analysis

Once a Gap has been identified, it is important to provide a preliminary analysis as to the criticality and priority of the gap and whether the analysis needs to be elevated to a higher level of authority. Based on the preliminary analysis, the decision may be to accept the gap, conduct more detailed analysis, establish risk mitigation activities to reduce the risk, or elevate the gap analysis to a higher level of authority. If the decision is to conduct more detailed analysis, then typical root cause analysis may be conducted to determine if there are broader consequences due to the identified Gap. Once all Gap Analysis is conducted, root cause identified, prioritization determined, and next actions defined, this information is appended to the Gap Identification information and this information passed on to the Gap Risk Mitigation step.

3.4.3. Gap Risk Mitigation

Once a Gap Analysis has been conducted responsible risk agents need to determine what the potential process/product risk is associated with the gap along with a prioritization and mitigation strategy. Risk resolution defines the activities to be conducted which in turn result in possible product/process changes, or in some cases changes to the Normative References which may take a very mature organization to accomplish without actually increasing the resulting risk and quality indicator gap measure. Potential risk options (combinations are possible) include:

- Risk avoidance: Changing or lowering requirements while still meeting the user's needs
- Risk control: Taking active steps to minimize risks
- Risk transfer: Reallocating design requirements to lower the risks
- Risk monitor: Watching and periodically reevaluating the risk for changes to the assigned risk parameters
- Risk acceptance: Acknowledgment of risk but deciding not to take any action

3.4.4. Gap Measurement Scales

It must be understood that Risk Indicator Gaps may be represented on nearly any of the measurement scales. This is important to understand because mathematical operations that are valid to be performed on the gap measures will depend on the measurement scale. The measurement scale information is illustrated in Table 3-1.

Table 3-1. Gap Measurement Scales and Descriptions

Nominal	Ordinal	Interval	Ratio
People or objects with the same scale value are the same on some attribute. The values of the scale have no ‘numeric’ meaning in the way that you usually think about numbers.	People or objects with a higher scale value have more of some attribute. The intervals between adjacent scale values are indeterminate. Scale assignment is by the property of “greater than,” “equal to,” or “less than.”	Intervals between adjacent scale values are equal with respect the attribute being measured. E.g., the difference between 8 and 9 is the same as the difference between 76 and 77.	There is a rational zero point for the scale. Ratios are equivalent, e.g., the ratio of 2 to 1 is the same as the ratio of 8 to 4.

For example, measurement is in a Ratio scale, then normal arithmetic operations of addition, subtraction, multiplication, division and the associated statistical measurement associated with mean, standard deviation and so forth can be computed. For an Interval scale, differences can be calculated, but one can not compute multiplication and division operations – and hence statistical measures that require such operations are invalid. For the Ordinal scale, one can not use any of the standard arithmetic operations, but ranking/ordering operations can be done. For Nominal scales, there is no numeric comparisons – only attribute characteristics that provide the concept of group membership.

These concepts can be easily misunderstood since many times Interval or Ordinal scale data is represented by the symbol that is also a numeric value (e.g. integers). However, just because one represents information such as below with numbers does not mean those number symbols also inherit numeric operations.

Symbol	Meaning
0	Very Low Risk
1	Low Risk
2	Medium Risk
3	High Risk
4	Very High Risk

Not only is this not a Ratio scale, it isn’t even an Interval scale (although it may look like one). This is more easily understood if the symbols used were VL, L, M, H, VH.

3.5. Risk Model Details

The Risk Model determines potential risk indicators. This is accomplished by using the gaps (vulnerabilities) from the Evaluation Model, knowledge about the cognitive system application environment scenarios (intended as well as unintended use that covers normal, abnormal, and hostile environment events), and knowledge about potential threats that might be able to exploit the vulnerabilities. This model representation is illustrated in Figure 3-5, and illustrated the connectivity of the Specification, Design, and Evaluation Models.

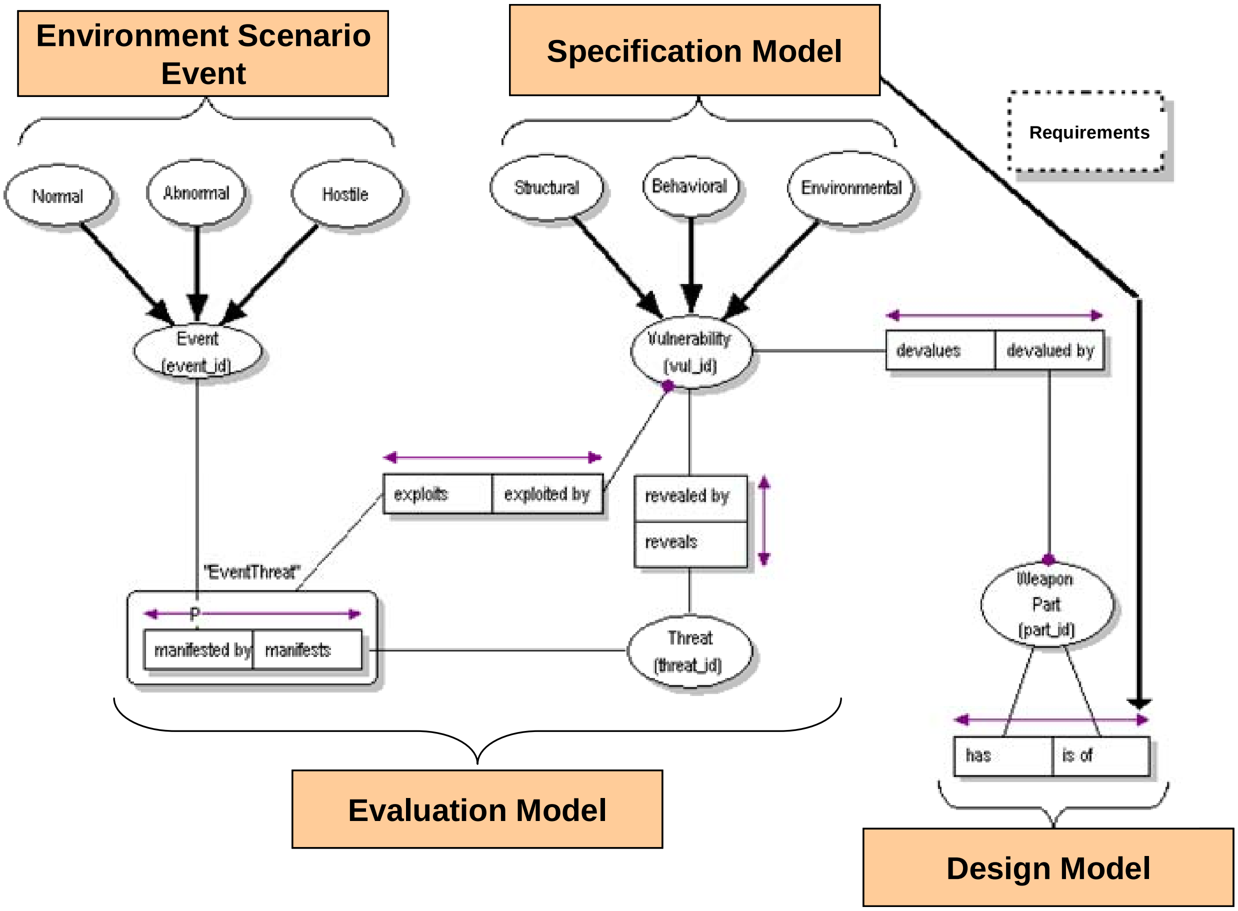


Figure 3-5. Cognitive System Conceptual Risk Model

For a specified environment scenario event, identified or potential cognitive system vulnerabilities and identified event threats the risk model provides:

(1) risk identification – identification of the possible threat/vulnerability pairing; uses gap identification;

(2) risk analysis – determination of the potential consequence/impact and likelihood of this identified risk as well as its prioritization among other identified risks; uses gap analysis information from the Evaluation Model as well as information on previous gaps of which the identified gap may be recurring;

(3) risk mitigation – determination of the strategy required to reduce the risk to an acceptable level; uses gap risk mitigation information from the Risk Model;

(4) risk management – tracking and management control of the identified risks across the full product life cycle model; monitors the vulnerabilities, risk mitigations, changing threats, and potential environmental scenario changes; instantiates continued Evaluation Model analysis throughout the product life cycle.

Vulnerabilities/gaps might be surety and/or ELS related. A gap might be an instance of a cognitive system requirement not satisfying a normative reference such as safety principles; or, perhaps a requirement is for use of a new technology that does not adhere to the safety principles. A gap might be between model instances, e.g., the design model instance does not implement the requirement to the specified acceptance criteria. The threat may be simply from the normal intended use of the cognitive system such that when the vulnerability (e.g., defect in the product) is encountered a person's safety is at risk. The threat may be from an unintended use scenario event where a person's personal information is obtained by someone without access authorization and the information was able to be compromised due to the lack of an adequate security authentication design and implementation. Consequence and Likelihood tables associated with the application can be defined early in a product life cycle and dictate the practice level that should be applied to reduce the risk to an acceptable level.

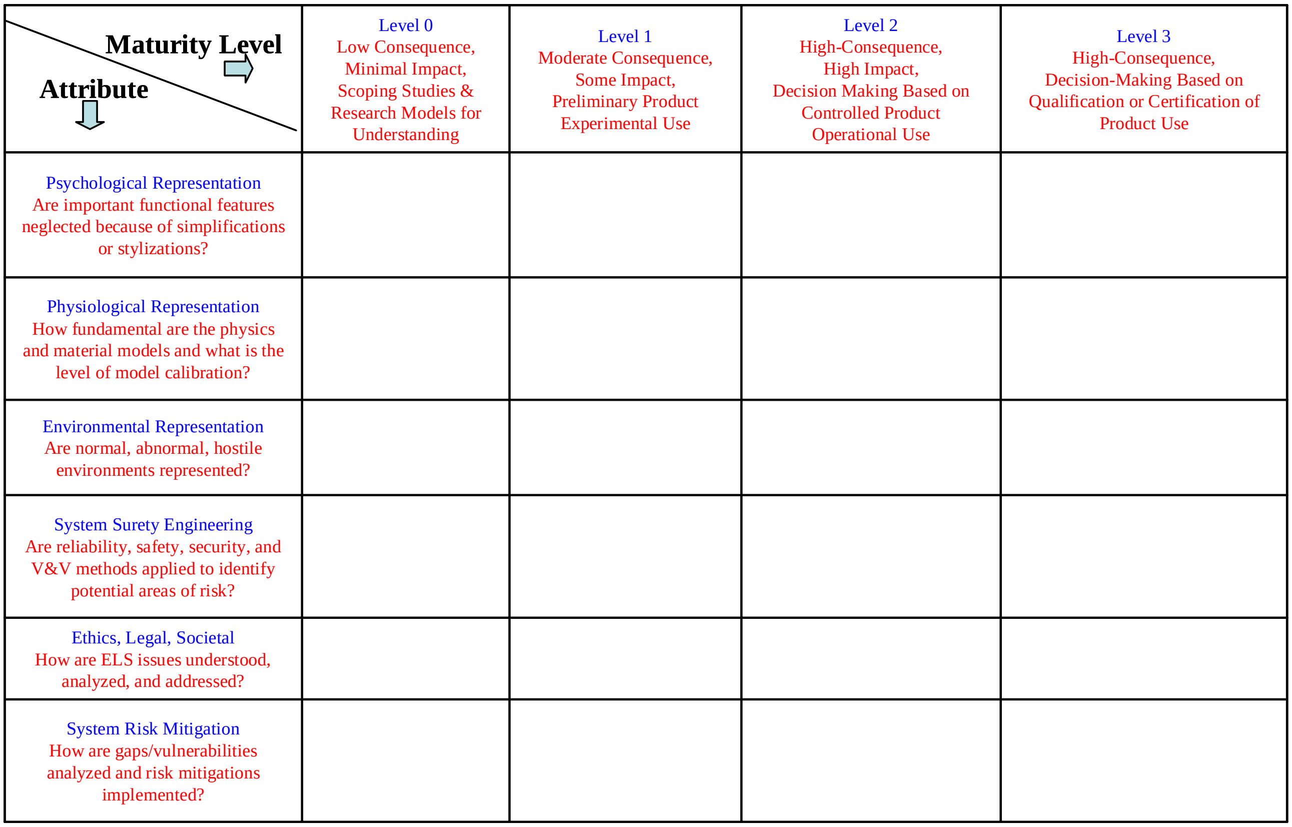
In addition, the Risk Model results will be an input to determine where the cognitive system maturity is in relation to the Cognitive System Maturity Model (CSMM) as described in the next section. If the existing maturity level is not what was expected in the CSMM for the intended application use, then it may be appropriate to develop an improvement strategy that incorporates the risk mitigation concepts.

3.6. Cognitive System Maturity Model Details

The Cognitive System Maturity Model (CSMM) provides fidelity characteristics that define how mature the cognitive system representation is from the perspective of risk aversion. The higher the targeted risk aversion, the more mature the cognitive representation needs to be. There are several neurological aspects of "representation" that factor into the fidelity considerations. In addition, there are several engineering surety and ELS aspects that also factor into the fidelity considerations. The fidelity/maturity is determined through application of a risk-informed decision analysis process that uses the results from the Risk Model to influence decision-makers who may have to use risk dimensions other than the technical and ELS dimensions in this framework. The fidelity level represents the "plausibility" of the cognitive system model representation of human cognitive processes.

CSMM is intended to help measure cognitive system fidelity progress, specify current cognitive model predictive capability, and help prioritize future cognitive system improvement. At this time, this model is just conceptual although the case study examples will provide some indication what the matrix cell content might contain. Completion of this conceptual CSMM is one recommendation from this research effort.

The conceptual CSMM is illustrated in the matrix of Figure 3-6a.



Maturity Level ⇨ Attribute ⇩	Level 0 Low Consequence, Minimal Impact, Scoping Studies & Research Models for Understanding	Level 1 Moderate Consequence, Some Impact, Preliminary Product Experimental Use	Level 2 High-Consequence, High Impact, Decision Making Based on Controlled Product Operational Use	Level 3 High-Consequence, Decision-Making Based on Qualification or Certification of Product Use
Psychological Representation Are important functional features neglected because of simplifications or stylizations?				
Physiological Representation How fundamental are the physics and material models and what is the level of model calibration?				
Environmental Representation Are normal, abnormal, hostile environments represented?				
System Surety Engineering Are reliability, safety, security, and V&V methods applied to identify potential areas of risk?				
Ethics, Legal, Societal How are ELS issues understood, analyzed, and addressed?				
System Risk Mitigation How are gaps/vulnerabilities analyzed and risk mitigations implemented?				

Figure 3-6a. Cognitive System Maturity Model

3.6.1.1. Maturity Levels

This CSMM includes four levels of maturity.

Level 0 represents typical research and very early prototype efforts whose failure or use would have low consequence with minimal impact. Projects/products in this level are characterized as Scoping Studies & Research Models for Understanding aspects of cognitive systems. The framework application is primarily for ensuring potential areas of risk are identified for consideration in future extensions of this work. Plausible representations are typically targeted to very specialized cognitive system maturity attributes without much representation of other areas.

Level 1 represents more sophisticated projects that typically result in a valuable prototype or early development of a useful cognitive system or product that is part of a cognitive system. Failure or use of the project results would have moderate consequence with some impact. Projects/products in this level are characterized as Preliminary Product Experimental Use. The framework application can now be used to make sure that surety and ELS engineering concerns are being addressed and that any operational production versions of the cognitive system product(s) will be designed to manage vulnerabilities to a target risk (hopefully an acceptable level of risk). Plausible representations still may be focused on one or more of the cognitive system maturity attributes, but the applicability of those attributes that are not a focus has been analyzed and potential risks (if any) identified.

Level 2 represents those cognitive system products that would have a high-consequence, high impact if the project failed or the product operational use failed (at least worst case). Projects/products at this level are characterized as Decision Making Based on Controlled Product Operational Use. In other words, the cognitive systems are typically products in the market place that may affect the user or some other persons in the manner that the cognitive systems are being used. Since the product is available, the potential impact of surety or ELS engineering failures could be high – cost and legal fees to the supplier, regulatory and other legal concerns, ethical ramifications for inadequate consideration of an individual or profile group privacy constraints, and even injury or death for individuals associated with the use of the cognitive system. The framework application can now be used to make sure that surety and ELS engineering concerns have been addressed and that operational production versions of the cognitive system product(s) are designed to manage vulnerabilities to a target risk (hopefully an acceptable level of risk). Potential areas of risk are identified for consideration in future extensions of this work. Plausible representations still may be focused on one or more of the cognitive system maturity attributes, but the applicability of those attributes that are not a focus has been analyzed and potential risks (if any) significantly reduced in the implemented product.

Level 3 is similar to Level 2 except there are regulatory requirements for certification/qualification of the cognitive system product in accordance with potentially high consequence use. Projects/products at this level are characterized as Decision Making Based on Qualification or Certification of Product Use. In other words, the cognitive systems are typically products in the market place that decision makers DEPEND on having gone through certification rigor with validation evidence for the qualified use of the cognitive system product in the specified application use. The potential impact of surety or ELS engineering failures is not only high in the sense of Level 2, but may impact a whole product line, an existing area of research, or a new product technology research and development effort. The framework application can now be used to its full extent with well-defined normative references, models, maturity level, and documented surety and ELS claims and arguments. Public acceptance is critical to the success of these cognitive system technologies and that acceptance would have been adequately vetted at this level. Applicable surety and ELS engineering concerns are documented and operational production versions of the cognitive system product(s) are designed to manage vulnerabilities to a target risk (which is an acceptable level of risk). Plausible representations still may be focused on one or more of the cognitive system maturity attributes, but the applicability of those attributes that are not a focus has been analyzed and potential risks (if any) are reduced to an acceptable level in the implemented product.

3.6.1.2. Maturity Attributes

The maturity attributes are simply the main elements of the Surety Engineering Cognitive System Framework – including any associated processes and practices necessary to implement the framework.

Psychological Representation: Are important psychological functional features neglected because of simplifications or stylizations? How representative of reality are the psychological models – requirements and design? Is there a purpose why some areas have a low fidelity, such as augmented cognitive function, or not important to the cognitive system?

Physiological Representation: How fundamental are the physics, chemistry, brain physiology, and material models and what is the level of model calibration? How representative of reality are the physiological models – requirements and design? Is there a purpose why some areas have a low fidelity – such as augmented connectivity representation function, or not important to the cognitive system?

Environmental Representation: Are normal, abnormal, hostile environments represented?

System Surety Engineering: Are reliability, safety, security, and V&V methods applied to identify potential areas of risk?

Ethics, Legal, Societal Issues: How are ELS issues identified, understood, analyzed, and addressed?

System Risk Mitigation: How are gaps/vulnerabilities analyzed and risk mitigations implemented?

A preliminary example of what a conceptual CSMM matrix would contain is illustrated in Figure 3-6b. The reference on the Predictive Capability Maturity Model (PCMM) for Computational Modeling and Simulation[12] provides a maturity model for verification and validation of science-based modeling and simulation. For cognitive systems, the PCMM model complements the overall surety engineering framework maturity model for V&V fidelity. Recent internal reports[13,14] have significantly enhanced the PCMM details. The PCMM will provide more depth to the V&V information illustrated in Figure 3-6b.

[12] W. Oberkampf, M. Pilch, T. Trucano, "Predictive Capability Maturity Model (PCMM) for Computational Modeling and Simulation, " SAND2007-5948, October 2007.

[13] M. Pilch, "PCMM 2nd Generation," internal SNL Word document, September 30, 2008.

[14] M. Pilch, T. Trucano, "PCMM Layout," internal PowerPoint presentation, October 27, 2008.

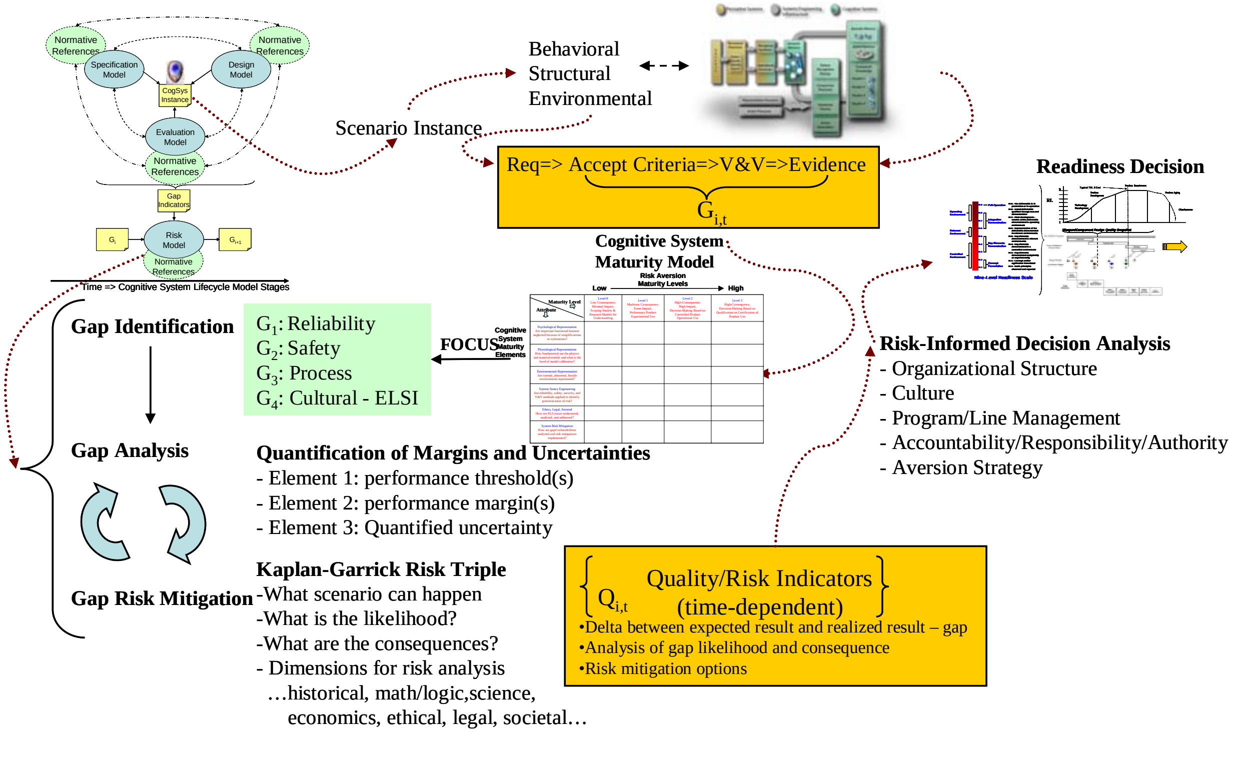
Maturity Level ⇨ **Attribute** ⇩	Level 0 Low Consequence, Minimal Impact, Scoping Studies & Research Models for Understanding	Level 1 Moderate Consequence, Some Impact, Preliminary Product Experimental Use	Level 2 High-Consequence, High Impact, Decision Making Based on Controlled Product Operational Use	Level 3 High-Consequence, Decision-Making Based on Qualification or Certification of Product Use
Psychological Representation Are important functional features neglected because of simplifications or stylizations?	• Judgment only • Little or no representational or behavioral fidelity for the cognitive system • Theory/Requirements are in development	• Significant simplification or stylization of the major components • Major elements are represented by behavioral models • Theory/requirements are understood and at least partially documented	• Limited simplification or stylization of major components • Significant simplification or stylization of other components • Models well defined for major components and some definition for minor components • Theory/Requirements are well-understood and documented	• Essentially no simplification or stylization of any components in the cognitive system • Detailed representation of all components and features “as built”, e.g., functions, interfaces, dependencies
Physiological Representation How fundamental are the physics and material models and what is the level of model calibration?	• Completely empirical model for the structural fidelity of the cognitive system • Few, if any, science-based models or mathematics to support the models • Theory/Requirements are in development	• Models dominated by calibration with Integral Effects Tests (IET) • All models are empirical with no basis for extrapolation to other applications • Theory/requirements are understood and at least partially documented	• Some calibration of models based on IETs • Significant calibration of models using Separate Effects Tests (SET) • Models are empirical with some basis for extrapolation to other applications • Theory/Requirements are well-understood and documented	• Minimal calibration of models based on IETs and SETs • Science-based models for all important physiological models • Important models have ability to extrapolate to other applications
Environmental Representation Are normal, abnormal, hostile environments represented?	• Judgment only • Limited normal scenarios • Theory/Requirements are in development	• Normal scenarios • Limited abnormal scenarios • Theory/requirements are understood and at least partially documented	• Normal scenarios • Abnormal scenarios • Limited hostile scenarios • Theory/Requirements are well-understood and documented	• Normal scenarios • Abnormal scenarios • Hostile scenarios
System Surety Engineering Are reliability, safety, security, and V&V methods applied to identify potential areas of risk?	• Judgment only • Numerical errors have an unknown or large effect on simulation results • Little of any attention to surety concerns, v&v, or risks – other than perhaps as required for legal compliance or future research/development plans • Theory/Requirements are in development	• Experimental safety analysis and mitigations in place • Experimental reliability of results are analyzed for critical parameters • Experimental Security/ Privacy issues are addressed • Some verification evidence available - sensitivity of numerical errors have been investigated on important System Response Quantities (SRQ) • Theory/requirements are understood and at least partially documented	• Experimental safety analysis and mitigations in place and verified • Operational safety analysis and mitigations are in place • Experimental reliability of results are estimated for critical parameters • Operational reliability of the system is estimated for critical parameters • Experimental Security/ Privacy issues are analyzed and implemented • Operational Security/Privacy issues are understood and mitigation mechanisms are implemented • Experimental and operational verification evidence available -	• Rigorous numerical errors have been estimated for all SRQs over the range of input parameters • Numerical errors due to UQ for important SRQs have been estimated • Experimental and Operational safety analysis and mitigations in place, verified, and validated • Operational safety analysis and mitigations are in place • Experimental reliability of results are estimated for critical parameters • Operational reliability of the system is estimated for critical parameters • Experimental Security/ Privacy issues are analyzed and implemented

			numerical errors have been quantitatively estimated on important SRQs; - Experimental validation evidence available - some numerical errors due to UQ have been estimated - Theory/Requirements are well-understood and documented	- Operational Security/Privacy issues are understood and mitigation mechanisms are implemented - Experimental and operational verification evidence available - numerical errors have been quantitatively estimated on important SRQs; - Experimental validation evidence available - some numerical errors due to UQ have been estimated
Ethics, Legal, Societal Issues How are ELS issues understood, analyzed, and addressed?	- Judgment only - Experimental concerns may be addressed to some extent-particularly for legal/contractual constraints - Individual privacy, safety protection for experimental research activity is qualitatively addressed per requirements - Product/research implications for the general public are unknown	- Product/process ELS system technical concerns related to privacy, safety, reliability have been identified and partially addressed with V&V evidence - Product/process ELS external public concerns are partially identified, but not specifically addressed - Product/process intended use and ELS issues have been identified and partially addressed. - Product/process unintended use and associated ELS issues have been partially identified, but not specifically addressed.	- Product/process ELS system technical concerns related to privacy, safety, reliability have been identified and addressed with adequate V&V and surety evidence - Product/process ELS external public concerns are identified and partially addressed - Product/process intended use and ELS issues have been identified and addressed through technical, management, public review. - Product/process unintended use and associated ELS issues have been identified, and partially addressed through technical reviews.	- Product/process ELS system technical concerns related to privacy, safety, reliability have been identified and addressed with V&V, surety, and QMU evidence - Product/process ELS external public concerns are identified and fully addressed through technical, management, and public review. - Product/process intended use and unintended ELS issues have been identified and addressed through technical, management, public review.
System Risk Mitigation How are gaps/vulnerabilities analyzed and risk mitigations implemented?	- Judgment only - Significant research/ scoping gaps are identified for future research	- Qualitative assessment of model accuracy with some important SRQ gaps and associated impacts identified - Large and/or unknown uncertainties in experimental evidence and tests - Top level technical risk areas identified - Some analysis of potential ELS risks completed	- Quantitative assessment of predictive accuracy for important SRQs with IETs and SETs - Some variabilities and uncertainties (V&U) are estimated for important SRQs - Major technical risk areas quantified and some mitigation strategies in place - ELS major risks identified addressed with the technical implementation	- Quantitative assessment of predictive accuracy for all important SRQs with IETs and SETs over the range of operating conditions of interest - All important V&Us are well characterized for all IETs and SETs - Major technical risk areas quantified and mitigation strategies in place - ELS major risks identified and technical implementations to reduce such risks have been verified and validated.

Figure 3-6b. Example of Conceptual CSMM Matrix

3.7. Risk-Informed Decision Process

The integration of the framework models is part of a risk-informed decision process. All of the elements of this process are illustrated in Figure 3-7. It is this full process that provides the time-dependent results from application of the framework to a specific cognitive system application – product or project.



Maturity Level / Attribute	Level 0 Low Consequence, Minimal Impact, Scoping Studies & Research Models for Understanding	Level 1 Moderate Consequence, Some Impact, Preliminary Product Experimental Use	Level 2 High-Consequence, High Impact, Decision Making Based on Controlled Product Operational Use	Level 3 High-Consequence, Decision-Making Based on Qualification or Certification of Product Use
Psychological Representation Are important functional features neglected because of simplifications or stylizations?				
Physiological Representation How fundamental are the physics and material models and what is the level of model calibration?				
Environmental Representation Are normal, abnormal, hostile environments represented?				
System Surety Engineering Are reliability, safety, security, and V&V methods applied to identify potential areas of risk?				
Ethics, Legal, Societal How are ELS issues understood, analyzed, and addressed?				
System Risk Mitigation How are gaps/vulnerabilities analyzed and risk mitigations implemented?				

G_1: Reliability
G_2: Safety
G_3: Process
G_4: Cultural - ELSI

Risk-Informed Decision Analysis
- Organizational Structure
- Culture
- Program/Line Management
- Accountability/Responsibility/Authority
- Aversion Strategy

Quantification of Margins and Uncertainties
- Element 1: performance threshold(s)
- Element 2: performance margin(s)
- Element 3: Quantified uncertainty

Kaplan-Garrick Risk Triple
-What scenario can happen
-What is the likelihood?
-What are the consequences?
- Dimensions for risk analysis
…historical, math/logic,science, economics, ethical, legal, societal…

$\left\{ Q_{i,t} \quad \text{Quality/Risk Indicators (time-dependent)} \right\}$
•Delta between expected result and realized result – gap
•Analysis of gap likelihood and consequence
•Risk mitigation options

Figure 3-7. Cognitive System Conceptual Risk-Informed Decision Process

This risk-informed decision process is an integrated discourse of the Risk Model application from Section 3.5. Once the framework is applied to a specific cognitive system instance (see Section 4) as in the upper left corner of Figure 3-7, the Risk Model can be applied. For a specific Scenario Instance, the Specification Model requirements are compared with the Design Model implementation using the Evaluation Model to determine any gaps in the requirements dimension: requirement acceptance criteria to V&V evidence. Such gaps are recorded in comparison with the CSMM and specific gaps with prioritized focus (such as the ones in the example green box) go through the Evaluation Model gap and risk mitigation analyses. It should be emphasized that this process is iterative and previous steps may be revisited depending on the results of any process step. Once the prioritized risk indicators are identified with risk mitigation options, the results are presented to decision makers who may consider other project, cultural, accountability risk dimensions have not been considered. This may involve internal technical teams, management teams, external peer review teams or even public discussion forums. Since this process is time-based, meaning it can occur throughout the project life cycle, it is possible to establish mechanisms early in the process that may prevent later undesirable life cycle events, like project/product critical failure.

4. PROCESS FOR APPLYING THE FRAMEWORK TO COGNITIVE SYSTEM IMPLEMENTATIONS

Application of the Surety Engineering Framework for Cognitive Systems is an iterative process over the project/application/product life cycle. The following general steps may be iterated as necessary depending on the life cycle stage.

4.1. Classify Cognitive System Application

Describe the cognitive system application with the conceptual framework concepts. Indicate which parts of the framework the implementation of interest is associated. This is a scoping exercise – ensure that intended and unintended scenario use cases are specified as applicable for the desired maturity level. Consider the normative references that are pertinent for the application, if any exist. This provides the relationship between the cognitive system application and existing standards, guidelines, state-of-the-art models and implementations. The desired maturity level is dependent on the life cycle stage of the project/application. The typical stages of a systems engineering life cycle model are illustrated in Table 4-1.

Table 4-1. Example of Systems Engineering Life Cycle Model

Life Cycle Phase	CSMM Characteristics/Level	Description
Stage 1 Early Concept Prototype Elements	**Maturity Level 0** Basic principles observed and reported Technology concept and/or application formulated.	Lowest level of technology readiness. Scientific research begins to be translated into applied research and development. Examples might include paper studies of a cognitive system technology's basic properties. Invention begins. Once basic principles are observed, practical applications can be invented. Applications are speculative, and there is no proof or detailed analysis to support the assumptions. Examples are limited to analytical studies.
Stage 2 Feasibility Study, Design Definition, Cost Study, Concept Selection	**Maturity Level 0/1** Analytical and experimental critical functions and/or characteristic proof of concept. Component and/or breadboard validation in laboratory environment.	Active research and development are initiated, which includes analytical and laboratory studies to physically validate analytical predictions of separate technology elements. Examples include components that are not yet integrated or representative. Basic technology components are integrated to establish that they will work together. This is a relatively low fidelity compared to the eventual system. Examples include integrating ad hoc hardware in the laboratory.
Stage 3 Development Engineering	**Maturity Level 1/2** Component and/or breadboard validation in relevant environment. System/subsystem model or prototype demonstration in a relevant environment.	Fidelity of breadboard technology increases significantly. The basic technology components are integrated with reasonably realistic supporting elements so they can be tested in a simulated environment. Examples include high-fidelity integration of components in a laboratory. Representative model or prototype system, which is well beyond that of concept stage. Breadboard version is tested in a relevant environment, which represents a major step up in the plausible representation of the cognitive systems. Examples include testing a prototype in a high-fidelity laboratory or simulated operational environment.

Production Engineering Prove-In	System prototype demonstration in an operational environment.	Prototype is near or at planned operational system. Represents a major step up from Stage 2 and requires demonstrating an actual system prototype in an operational environment such as an aircraft, vehicle or in space. Examples include testing the prototype in a testbed operational environment representative of the operational environment. In many commercial developments this is the first delivery use depending on the criticality of potential failure in the field.
Stage 4 Production Engineering Certification	**Maturity Level 2/3** Actual system completed and "qualified / certified" through validation test and demonstration.	Technology has been proven to work in its final form and under expected conditions. In almost all cases, this stage is the end of true system development. Examples include developmental test and evaluation of the system in its intended environment to validate it meets design specifications. Depending on the certification rigor, this maturity level might be a 2 or 3.
Stage 5 Commercialization Full Scale Production Operational Use Support	**Maturity Level 2/3** Actual system operationally used through successful mission operations.	Actual application of the technology in its final form and under mission conditions such as those encountered in operational test and evaluation. In almost all cases, this is the end of the operational use and modification update/support aspect of the system life cycle. Depending on the certification rigor, this maturity level might be a 2 or 3.
Stage 6 Retirement	**Maturity Level N/A** Application system capabilities are not adequate for mission success or can be replaced by other systems	The retirement of an application involves the "de-certification" for use and appropriate archiving of important records as to the operational use, maintenance support, and lessons learned from the system. Although the maturity level is indicated as "Not Applicable – N/A", the retirement process will be completed easier the more mature the product is.

4.2. Derive Cognitive System Implementation Models

Define the steps used to derive instances of the various models specific to the application of interest. In particular, derive the application requirements and acceptance criteria, and related normative references. Strongly consider, depending on the desired maturity level, what surety and ELS requirements are relevant. Develop the design architecture and indicate what part of the framework conceptual Design Model applies. Indicate any use of existing normative references for elements of the Design Model. In particular, identify how surety and ELS requirements are being implemented in the design architecture. Define the verification and validation approach and how the application design will be evaluated per the Evaluation Model methods and techniques. Describe the use of the conceptual framework models to facilitate use of surety engineering and ELS engineering standard practices; lessons learned from previous efforts; existing architectures/designs/characterizations and evaluation methods and tools that can be applied to facilitate the conceptual research stage and/or other life cycle stages. Emphasize how the available checklists were used, if at all. Ensure that intended and unintended environment scenarios are derived to support the risk model and early identification of potential risk areas.

4.3. Determine Potential Areas of Risk

Apply the normative references, lessons learned, environment scenarios, and other aspects of the risk-informed decision model depending on the life cycle activity of concern and the desired cognitive system model maturity level. Be sure to address surety risks as well as ELS risks. Prioritize risks and determine how surety methods are being or could be applied to reduce surety technical risks as well as potential ELS risks. Identify areas of potential unknowns that can be addressed during the research/development.

4.4. Conduct Implementation Activities

These activities may be research for understanding and/or may be more formal product development for operational use. Incorporate surety methods and techniques in both the engineering and assurance of the implementation. Conduct design characterization and use evaluation methods to establish if there are potential process/product gaps – both technical and ELS. Address risk indicators through the risk model scenario for threat/vulnerability across behavioral, structural, and environmental requirements. Conduct risk analyses and management using the risk-informed decision model appropriate to the implementation goals.

4.5. Provide Implementation Results

Depending on the implementation goals, the results may range from research to final product release. Internal and external review of the implementation results and identified areas of risk may require internal organization and external SME peer reviews, as well as scenario-based collaborations with regulators, customers, users, public, and other stakeholders. Research results can be published for other applications of the framework. Monitoring can be continued for any product implementations to detect potential problems or additional use possibilities.

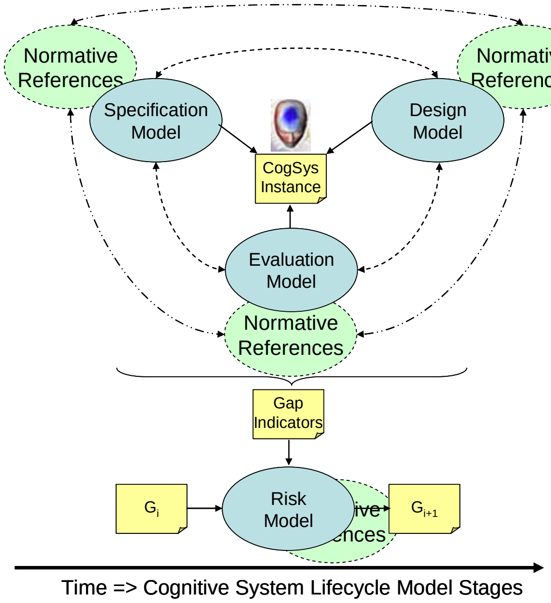
4.6. Conceptual Model Framework Feedback

Every implementation instance will provide lessons learned and updates to the framework’s normative information and models. In this way the framework itself can evolve as more information is obtained, normative references are improved, conceptual models are improved, and processes and tools for implementation are developed.

5. CASE STUDY: NON-INVASIVE COGNITIVE STATE DETECTION

This section provides a short case study of an existing Sandia research project to illustrate application of the Surety Framework. This case study is not meant to be comprehensive, but only an example to illustrate some of the ideas in a more practical application setting. The research project is analyzed as a cognitive system instance, addressing each of the framework models, associated normative references and gaps/vulnerabilities, and an overall appraisal of the surety and ELS risks that might be manifested by this cognitive system instances.

The Non-Invasive Cognitive State Detection LDRD project, funded in FY2005, was intended to explore the creation of a computer workstation capable of integrating a number of non-invasive sensors, as well as Sandia's cognitive modeling capabilities, for the purpose of extracting a user's current cognitive state. This project resulted in the ability to use existing models of users or create new user models to provide real-time system adaptation to the user. The ability to automatically capture a user's cognitive state allows for the development of cognitive systems that are adaptable to an individual user and allows for the evolution of more accurate models of the individual.



5.1. Applying the Surety Framework to a Research Scenario

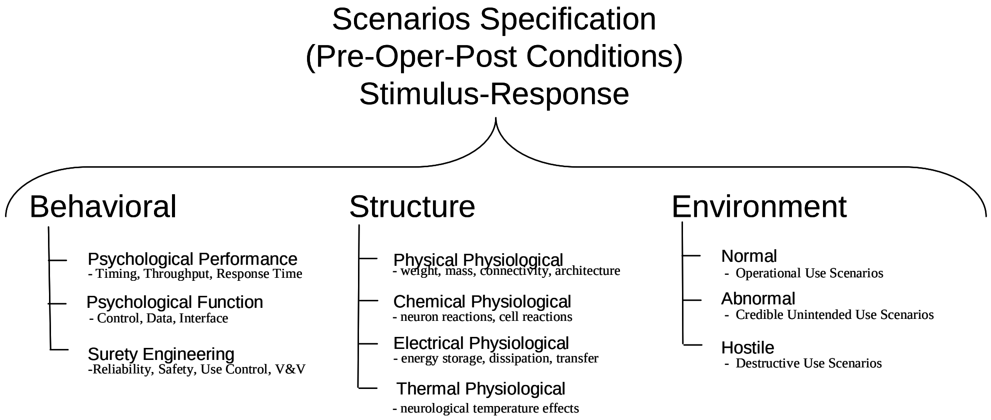
For this case study, the Surety Framework is applied to the Non-Invasive Cognitive State Detection research project. As a research project, the Surety Framework is applied in the context of *research* in the Early Concept stage of a program's lifecycle. A research project will typically have a large quantity of gap indicators and the associated Cognitive System Maturity Matrix is expected to be at a lower level.

The actual Integrated Workstation prototype developed for the Non-Invasive Cognitive State Detection project serves as the cognitive system instance of the framework. This prototype is currently in use by Sandia to define and further refine the specifications and intended uses for future applications.

Each of the framework's models is detailed in the following sections. The details for the models were extracted from the available research documentation as well as personal interviews with the project's Principle Investigator.

5.2. Specification Model

The Surety Framework's Specification Model considers the generic requirements with a focus on the behavioral, structural and environmental scenarios which could be applied. The Behavioral Model serves as the psychological basis executing cognitive functions consistent with plausible psychological models of how humans think. The Structural Model serves as the physiological basis executing cognitive functions consistent with plausible physical functioning of the human brain. Finally, the Environmental Model considers the plausible environments for the Behavioral and Structural Models including normal, abnormal, and hostile operational scenarios.


Scenarios Specification
(Pre-Oper-Post Conditions)
Stimulus-Response
Behavioral
Psychological Performance
- Timing, Throughput, Response Time
Psychological Function
- Control, Data, Interface
Surety Engineering
-Reliability, Safety, Use Control, V&V
Structure
Physical Physiological
- weight, mass, connectivity, architecture
Chemical Physiological
- neuron reactions, cell reactions
Electrical Physiological
- energy storage, dissipation, transfer
Thermal Physiological
- neurological temperature effects
Environment
Normal
- Operational Use Scenarios
Abnormal
- Credible Unintended Use Scenarios
Hostile
- Destructive Use Scenarios

5.2.1. Requirements

Explore the creation of a computer workstation capable of integrating a number of non-invasive sensors, as well as Sandia's cognitive modeling capabilities, in order to extract a potentially plausible representation of a user's current cognitive state. The goal is to combine measures of various aspects of cognitive state (i.e., attention, cognitive load, cognitive engagement, and emotional response) into a single system and demonstrate how these measures might be used to create adaptive systems capable of enhancing cognition.

5.2.2. Scenarios

<u>Behavioral</u>

The psychological basis of this work is limited to a selection of measurements which capture the cognitive state of an individual: attention, cognitive load, cognitive engagement, and emotional response. This selection was influenced based on a literature survey pertaining to human cognitive state and/or cognitive load. Further psychological studies are desired to capture the meaning and intent surrounding the workstation operator's behavior; however, this work has been delayed until further funding can be obtained.

Sandia's Human Subjects Board (HSB) review and approval was obtained prior to commencing this work. Also, the corporate ES&H policy was followed within the laboratory setting and use of equipment. A standard computer workstation with integrated sensors capable of collecting a variety of different measures of a user's cognitive state was the primary setup. The instruments were calibrated and operated per the manufacturer's directions. No additional use controls were added to this prototype demonstration beyond the typical password securities that existed for the workstation.

Structural

A review of existing approaches revealed that the four aspects of cognitive state selected to be measured for this project (attention, cognitive load, cognitive engagement, and emotional response) could be done using non-invasive means. The non-invasive physiological measurements were collected using the following sensory devices:

- remote eye tracker for attention capture
- posture sensing chair for cognitive engagement detection
- speech to text transcription technology for attention capture
- mouse and keyboard event tracking for measure of cognitive load
- pressure sensitive mouse to measure cognitive load and emotional response
- thermal imaging for measuring emotional response from operator’s face

Environmental

Normal Operating Scenarios

The normal use environment selected for the Integrated Workstation was a standard computer laboratory workstation with a researcher serving as the representative workstation operator. The workstation included instruments to collect sensor data to enhance the fidelity of a computer operator’s cognitive model as well as improve the human-operator interface by adapting to the operator’s dynamic cognitive state. To make the system plausible for deployment in the future, it was decided that only non-invasive sensors could be used to capture measures of cognitive state.

This environment fit a broad range of plausible problem domains. It also allowed for greater ease in monitoring the individual, since he/she is fairly stationary. Additionally, this environment is expected to allow for an easy transition to other important problem domains, such as the cockpit of an airplane, or the driver’s seat of a car.

Abnormal Operating Scenarios

Unintended case scenarios have been considered for this project; however, no attempts have been made at this time to prevent abnormal conditions from occurring. These unintended case scenarios include:

- concluding cognitive state from data results (Note: no psychological behavior has been correlated to the results; thus, such conclusions can not be valid)
- judging operator’s cognitive state (see note above)

Hostile Operating Scenarios

No hostile environments have been considered

Potential Consequences

Consequences emerge when vulnerabilities are associated with one or more threats. The vulnerabilities and threats identified for the project specific to the stated specification requirements are listed in the Table 5-1 below. The likelihood, impacts, and any mitigation plans for each consequence listed here are addressed in the Risk Model section.

Table 5-1. Specification Model Vulnerabilities, Threats, and Potential Consequences

Vulnerability	+	Threat	=	Potential Consequence
Researchers can be harmed by lab equipment	+	Workstation components (PC and sensory equipment) are unsafe to use	=	**Researchers are physically harmed**
	+	Workstation components (PC and sensory equipment) are used incorrectly	=	**Researchers are physically harmed**
Researchers will be allowing physiological data to be measured	+	Data collected from physiological measurements are not safeguarded	=	**Researchers' personal private information is compromised**
Data pertaining to researchers' cognitive state will be measured	+	Conclusions are made about cognitive state based on data collected from physiological measurements (not the goal of this research!)	=	**Unfair/inaccurate judgments can be made about researchers' cognitive state**
Project research relies on LDRD funds	+	LDRD funding is cut or reduced	=	**Project goals are not achieved**
			=	**Research is not performed at Sandia**

Normative References

Gazzaniga, Michael S., Ivry, Richard B., Mangun, George R., "Cognitive Neuroscience, Third Edition," W. W. Norton & Company, 2008.

References

Balaban, Carey D., Jarad Prinkey, Greg Frank, and Mark Redfern. "Automatic Event Structure Parsing for Context Modeling: A Role for Postural Orienting Responses." University of Pittsburgh.

Balaban, C. D., Cohn, J., Redfern, M. S., Prinkey, J., Stripling, R., Hoffer, M. 2004. Postural control as a probe for cognitive state: Exploiting human information processing to enhance performance. *International Journal of Human-Computer Interaction*. 17(2), 275-286.

Balaban, Carey D., Jarad Prinkey, Greg Frank, and Mark Redfern. "Postural measurements seated as gauges of cognitive engagement." University of Pittsburgh.

Forsythe, J. Chris, Bernard, Michael, L., and Goldsmith, Timothy E., editors, "Cognitive Systems: Human Cognitive Models in Systems Design," Lawrence Erlbaum Associates, Mahwah, New Jersey, 2006.

Crosby, Martha E., and David N. Chin. "Real Time Assessment of Cognitive Load for Adaptive Training." ONR Workshop. University of Hawaii at Manoa, Honolulu.

Ikehara, C. S., Crosby, M. E. 2005. Assessing cognitive load with physiological sensors. *Proceedings of the 38th Hawaii International Conference on System Sciences*.

Maylor, Elizabeth A., Sue Allison, and Alan M. Wing. “Effects of spatial and nonspatial cognitive activity on postural stability.” British Journal of Psychology 92 (2001): 319-38.

McClain, Jonathan T. “An Integrated Workstation for Determining a User’s Cognitive State”, SAND2005-3947 C.

McClain, Jonathan T. “An Integrated Workstation for Determining a User’s Cognitive State”, SAND2005-6102 P.

Pellecchia, Geraldine L. “Postural sway increases with attentional demands of concurrent cognitive task.” Gait and Posture 18 (2003): 29-34.

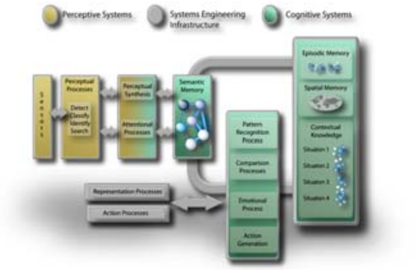
Qi, Yuan, Carson Reynolds, Rosalind W. Picard. “The Bayes Point Machine for Computer-User Frustration Detection via PressureMouse.” Massachusetts Institute of Technology.

Reynolds, Carson J. “The Sensing and Measurement of Frustration with Computers.” Massachusetts Institute of Technology (2001).

Sharma, Rajeev, Carey Baraban, and Keith Brendly. “Toward Non-intrusive Multimodal Emotional State Monitoring.” DARPA AugCog PI Meeting. Advanced Interfaces, Inc., Orlando. Jan. 2004.

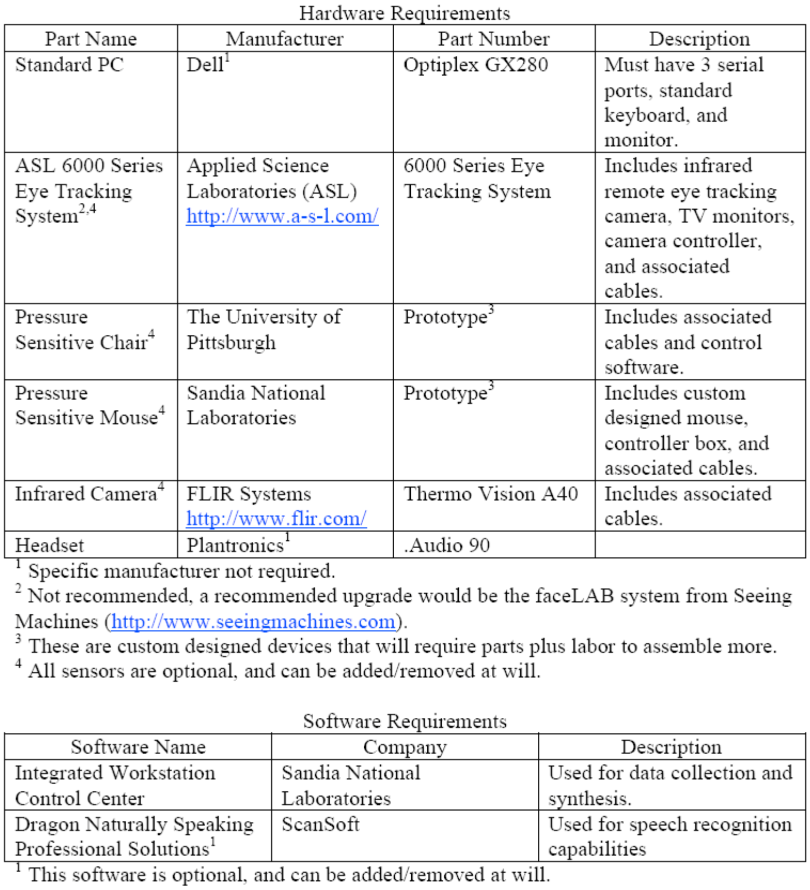
Skocypec, Russell D., et.al. 2006. “Next Generation Intelligent Systems Grand Challenge LDRD Final Report.” SAND2006-2506 pp. 64-73.

5.3. Design Model



The Integrated Workstation was designed to demonstrate how a workstation operator’s cognition could be augmented by applying the Sandia attention capture component to filter information based on the user’s cognitive state in real time. This application combines a textual model of the individual derived from previously gathered data with a real time measure of attention derived through use of an eye tracker. By comparing new documents that the user is attending to with the user’s model, the system is able to derive a real time measure of cognitive state. The application can then use this information to compare it to other textual data streams (i.e. instant messaging sessions) in order to draw attention to data relevant to the user’s current cognitive state. In this way, the Integrated Workstation can act as an information filter for users, drawing the user’s attention to information that is only relevant to the task at hand.

The list of the hardware and software components comprising the Integrated Workstation are listed in Figure 5-1.



Hardware Requirements

Part Name	Manufacturer	Part Number	Description
Standard PC	Dell[1]	Optiplex GX280	Must have 3 serial ports, standard keyboard, and monitor.
ASL 6000 Series Eye Tracking System[2,4]	Applied Science Laboratories (ASL) http://www.a-s-l.com/	6000 Series Eye Tracking System	Includes infrared remote eye tracking camera, TV monitors, camera controller, and associated cables.
Pressure Sensitive Chair[4]	The University of Pittsburgh	Prototype[3]	Includes associated cables and control software.
Pressure Sensitive Mouse[4]	Sandia National Laboratories	Prototype[3]	Includes custom designed mouse, controller box, and associated cables.
Infrared Camera[4]	FLIR Systems http://www.flir.com/	Thermo Vision A40	Includes associated cables.
Headset	Plantronics[1]	.Audio 90	

[1] Specific manufacturer not required.

[2] Not recommended, a recommended upgrade would be the faceLAB system from Seeing Machines (http://www.seeingmachines.com).

[3] These are custom designed devices that will require parts plus labor to assemble more.

[4] All sensors are optional, and can be added/removed at will.

Software Requirements

Software Name	Company	Description
Integrated Workstation Control Center	Sandia National Laboratories	Used for data collection and synthesis.
Dragon Naturally Speaking Professional Solutions[1]	ScanSoft	Used for speech recognition capabilities

[1] This software is optional, and can be added/removed at will.

Figure 5-1. Case Study Integrated Workstation Hardware & Software Components

Potential Consequences

Consequences emerge when vulnerabilities are associated with one or more threats. The vulnerabilities and threats identified for the project specific to the design model information is listed in the Table 5-2 below. The likelihood, impacts, and any mitigation plans for each consequence listed here are addressed in the Risk Model section.

Table 5-2. Design Model Vulnerabilities, Threats, and Potential Consequences

Vulnerability	+	Threat	=	Potential Consequence
Integrating suite of sensors with the workstation has not been done before	+	One or more sensors do not integrate with the workstation	=	**Project goals are not achieved**
Extracting user's cognitive state has not been done before	+	Cognitive state data can not be collected from measurements	=	**Project goals are not achieved**

Normative References

ASL 6000 Series Eye Tracking System. http://www.a-s-l.com/.

Dell Computer.

FLIR Systems. http://www.flir.com/.

Kandel, Eric R., Schwartz, James H., and Jessell, Thomas M., (editors): Principles of Neural Science, 4th ed. McGraw-Hill; 2000.

ScanSoft. Dragon Naturally Speaking Professional Solutions.

References

St. John, M., Kobus, D.A. & Morrison, J.G. (2003). DARPA Augmented Cognition Technical Integration Experiment (TIE).Technical Report 1905, SPAWARS Systems Center, San Diego, CA

5.4. Evaluation Model

The evaluation of this project involves assessing whether or not the requirement goals were met. This assessment was performed by the Principle Investigator's observation of the performance of the Integrated Workstation. Gaps are identified where the expectations of the goals were not met or could be improved upon, as indicated in Table 5-3.

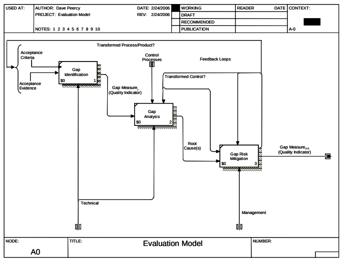


Table 5-3. Case Study Gaps and Improvement Opportunities

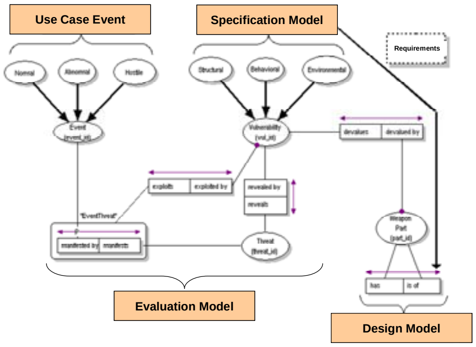
Requirement	Observation	Gap
Integrate sensors with workstation	All sensors integrated with workstation.	Validation and verification of the viability of each measure was not extensive. All sensors integrated successfully with the workstation. Additional measurements identified for future work are listed as follows: - expand the customized mouse to include temperature and Galvanic Skin Response sensors - voice stress analysis to infer emotional response - algorithm development for the thermal camera to extract various biometric measures from facial thermal images. - Measure of respiration might be captured by measuring temperature changes around the mouth and nose regions of the face - Measure of nose temperature has recently been linked with cognitive load
Integrate Sandia's cognitive modeling with workstation	Cognitive model integrated with workstation	No gaps
Extract user's cognitive state	Data measurements of the subject's cognitive state were collected using the sensors available.	Additional work might consider combining disparate measures into a single measure of cognitive state. However, it has not been determined that combining measures exceeds the performance of each measure alone. No conclusions of the user's cognitive state were definitively made. Further work should consider how the data collected can be associated with the psychological literature.

Potential Consequences

The evaluation phase involved the Principle Investigator using the Integrated Workstation as an operator and making observations about the data collected and its performance. No risks in addition to those identified in the Specification and Design Models were identified.

5.5. Risk and Maturity Models

The risks for each of the models are compiled in the table below. A probability index of high, medium or low has been assigned to each threat. This probability index can be used to determine how vital a mitigation plan might be for a project. In general, threats with a high probability should have a mitigation plan in place based on criteria of acceptance.



If no adequate mitigation plan exists, then a threat has the potential to exploit a vulnerability thus realizing the associated risk. Risks without a mitigation plan expose gaps in the surety of the project. For this Integrated Workstation, two gaps exist relating to project funding. While there are basic actions a Principle Investigator can perform to avoid instigating funding cuts (i.e., showing work is being performed) the decision to cut and/or reduce funds is typically beyond his/her control. Therefore, the risks relating to funding cuts existed as acceptable gaps with low probability that they would occur. Note also the gaps related to the technical risks. It is expected that a research project will have such gaps as without risk it is unlikely to be defined as research.

The mitigation plans and all gap indicators should be reviewed periodically as the project transitions to subsequent lifecycle phases. While these mitigation plans and gaps might be acceptable for this research project within a laboratory environment, it is unlikely they will be acceptable as the Integrated Workstation evolves into subsequent phases.

Cognitive System Maturity Matrix

The Cognitive System Maturity Matrix provides an at-a-glance view of the development level of the fidelity of the Integrated Workstation. Overall, the Integrated Workstation is clearly in the research stage where understanding of the concepts is the focus. There is limited attention to the consideration of an array of environments, surety, risks, and any ethical, legal, and social implications. The associated Cognitive System Maturity Matrix for this Case Study is illustrated in Table 5-5.

Table 5-4. Case Study Risk Model Summary

Vulnerability	Threat	Likelihood	Potential Consequence	Impact	Current Practices	Gap?	Mitigation Plan
Researchers can be harmed by lab equipment	Workstation components (PC and sensory equipment) are unsafe to use	Low	**Researchers are physically harmed**	High	ES&H Policies	No	
	Workstation components (PC and sensory equipment) are used incorrectly	Low	**Researchers are physically harmed**	High	ES&H Policies	No	
Researchers will be allowing physiological data to be measured	Data collected from physiological measurements are not safeguarded	Low	**Researchers' personal private information is compromised**	Medium	Access Control using PC password	No	
Data pertaining to researchers' cognitive state will be measured	Conclusions are made about cognitive state based on data collected from physiological measurements (not the goal of this research!)	Low	**Unfair/inaccurate judgments can be made about researchers' cognitive state**	Medium	Clearly communicate project objectives	No	
Integrating suite of sensors with the workstation has not been done before	One or more sensors do not integrate with the workstation	Medium	**Project goals are not achieved**	Medium	-	**Yes**	
Extracting user's cognitive state has not been done before	Cognitive state data can not be collected from measurements	High	**Project goals are not achieved**	Medium	-	**Yes**	
Project research relies on LDRD funds	LDRD funding is cut or reduced	Low	**Project goals are not achieved**	Medium	Execute project communicating progress to management	**Yes**	
			Research is not performed at Sandia	Medium	Execute project communicating progress to management	**Yes**	

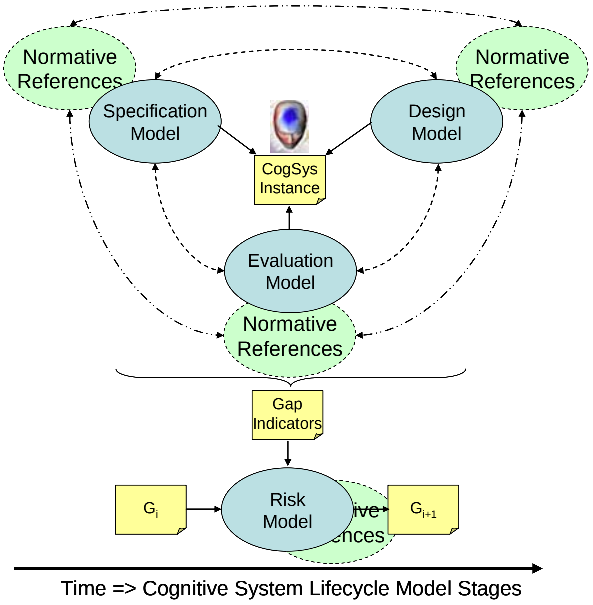
Table 5-5. Cognitive System Maturity Matrix

Fidelity Level Attribute	**Level 0** **Low Consequence, Minimal Impact Scoping Studies & Research Models for Understanding**	**Level 1** **Moderate Consequence, Some Impact Preliminary Product Experimental Use**	**Level 2** **High Consequence**	**Level 3** **High Consequence, Qualified/Certified**
Psychological Representation Fidelity	Basis for measurement of cognitive measure.			
Physiological Representation Fidelity		Physiological measurements using calibrated test equipment. No basis for extrapolation to other applications.		
Environmental Representation Fidelity	Normal environment defined. No abnormal scenarios identified. No hostile scenarios identified.			
System Surety Engineering Fidelity	Limited to laboratory environment.			
System Risk Mitigation Fidelity	Limited to research project's goals.			
Ethical, Legal and Societal Implications	Limited discussions. Misuse – obvious alternate uses were identified. Data results could be misconstrued to judge a workstation operator's performance. Data results between operators could be unfairly compared. Abuse – if this technology was misused, a supervisor could unfairly prohibit promotions or even demote operators.			

6. CASE STUDY: LONG-RANGE IRIS RECOGNITION IN NONIDEAL CONDITIONS

This section provides another short case study of an existing Sandia research project to illustrate application of the Surety Framework. This case study is not meant to be comprehensive, but only an example to illustrate some of the ideas in a more practical application setting. The research project is analyzed as a cognitive system instance, addressing each of the framework models, associated normative references and gaps/vulnerabilities, and an overall appraisal of the surety and ELS risks that might be manifested by this cognitive system instances.

The Long-Range Iris Recognition in Nonideal Conditions LDRD project, funded in FY2009, aims to create a high-accuracy, high-throughput iris recognition system that works with subjects and multi-meter distances. The project will explore new research and development in adaptive optics and software algorithms for iris recognition in nonideal conditions. The integrated system will provide a high-accuracy, high-throughput, multi-meter distance iris recognition of both cooperative and uncooperative subjects.



6.1. Applying the Surety Framework

For this case study, the Surety Framework is applied to the Long-Range Iris Recognition in Nonideal Conditions research project while in the early research phase. Applying the framework in the early phases of a research project is expected to identify a more comprehensive set of potential gaps that can be addressed early and mitigated as necessary; more effectively communicate the risk level and maturity; and capture the issues/concerns that could be transitioned beyond research phase.

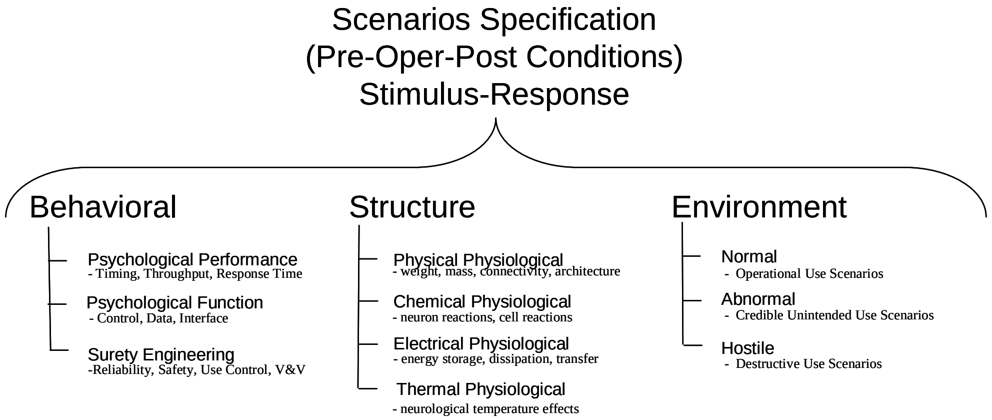
As a research project, the Surety Framework is applied at the Early Concept stage of a program's lifecycle. A research project typically has a large quantity of gap indicators and the associated Cognitive System Maturity Matrix is expected to be at a low maturity index level.

The actual Iris Recognition prototype that will be developed for the Long-Range Iris Recognition in Nonideal Conditions project will eventually become the cognitive system instance of this framework. This prototype will advance in three stages: 3-5 meter standoff iris recognizer for cooperative subjects; 10+ meter standoff iris recognizer for cooperative and uncooperative subjects; and a customer-specific design.

Each of the framework's models is detailed in the following sections. The details for the models were extracted from the available research documentation as well as personal interviews with the project's Principle Investigator.

6.2. Specification Model

The Surety Framework's Specification Model considers the generic requirements with a focus on the behavioral, structural, and environmental scenarios that could be applied. The Behavioral Model serves as the psychological basis executing cognitive functions consistent with plausible psychological models of how humans think. The Structural Model serves as the physiological basis executing cognitive functions consistent with plausible physical functioning of the human brain. Finally, the Environmental Model considers the plausible environments for the Behavioral and Structural Models including normal, abnormal, and hostile operational scenarios.


Scenarios Specification
(Pre-Oper-Post Conditions)
Stimulus-Response
Behavioral
Psychological Performance
- Timing, Throughput, Response Time
Psychological Function
- Control, Data, Interface
Surety Engineering
-Reliability, Safety, Use Control, V&V
Structure
Physical Physiological
- weight, mass, connectivity, architecture
Chemical Physiological
- neuron reactions, cell reactions
Electrical Physiological
- energy storage, dissipation, transfer
Thermal Physiological
- neurological temperature effects
Environment
Normal
- Operational Use Scenarios
Abnormal
- Credible Unintended Use Scenarios
Hostile
- Destructive Use Scenarios

Requirements

Create an iris-recognition system for cooperative and uncooperative subjects at multi-meter distances.

Phase I	3-5 meter standoff iris recognizer for cooperative subjects
Phase II	10+ meter standoff iris recognizer for cooperative and uncooperative subjects
Phase III	customer specific design

<u>Behavioral</u>

Some psychological phenomena might be considered to determine how uncooperative subjects might behave in order to avoid detection.

<u>Structural</u>

Noninvasive physiological measurements will be taken of human subjects' irises using Sandia's adaptive optics imaging system.

Sandia's Internal Review Board (IRB) will review and approve the procedures that will be used to protect human subjects as physiological measurements are performed and data is collected. Sandia's IRB will ensure adequate protections are in place for the human subjects and privacy concerns are addressed.

Sandia’s corporate ES&H policies will be followed to ensure safety within the laboratory environment including use of equipment.

<u>Environmental</u>

Normal Operating Scenarios

The Iris Recognition prototype is intended to be used on both cooperative and uncooperative subjects for the following relevant missions:

- Immigration and Customs Enforcement for two-factor verification using passport photos
- Transportation Security Administration for airport screening
- Securing Government Assets for watch-list identification and access verification
- Force Protection Malicious Intent for remote identification of intruders

Abnormal Operating Scenarios

None identified.

Hostile Operating Scenario

None identified.

Potential Consequences

Vulnerabilities associated with a threat pose risks. The vulnerabilities and threats identified for the project specific to the specification requirements are listed in Table 6-1 below. The likelihood, impacts, and any mitigation plans for each consequence listed here are addressed in the Risk Model section.

Table 6-1. Specification Model Vulnerabilities, Threats, and Potential Consequences

Vulnerability	+	Threat	=	Risk
Researchers and human subjects will be using or exposed to lab equipment	+	Laboratory equipment is unsafe to use	=	**Researchers or human subjects are physically harmed**
	+	Laboratory equipment is used incorrectly	=	**Researchers or human subjects are physically harmed**
Human subjects’ iris scans will be collected	+	Data collected from physiological measurements are not safeguarded	=	**Human subjects’ privacy might be compromised**
Project research relies on LDRD funds	+	LDRD funding is cut or reduced	=	**Research goals not achieved**
			=	**Research is not performed at Sandia**

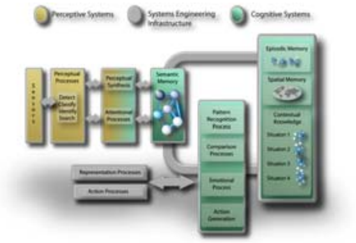
Normative References

National Institute of Standards and Technology

References

Dixon, Kevin R. “Iris Overview”, 2008.

6.3. Design Model



The design of the Iris Recognition prototype will address both the hardware and software aspects of multi-meter iris recognition for both cooperative and uncooperative subjects in non-ideal conditions. On the software side, we will research iris-recognition algorithms for non-ideal lighting conditions, face-pose reconstruction, non-orthogonal iris segmentation, and fusion of iris recognition and face recognition using quality measures. These algorithms will mitigate against subjects who are actively avoiding identification by looking away from likely imaging systems, wearing shadow-inducing hats, etc. On the hardware side, we will pursue hardware that is specially tailored toward identification and verification of both cooperative and uncooperative subjects at multi-meter distances. The imaging system will be based on Sandia’s expertise in foveated-zoom adaptive optics. These systems have a wide field of view and a narrow (zoomed) field of view that is steered by a micro-mirror, which can foveate anywhere within the wide field of view within about 10 milliseconds.

We will also pursue special-purpose lighting to determine the optimal time to acquire a subject’s iris. For example, we will investigate directional light emitters and detectors to automatically sense eyeball retroreflection (“redeye” or “cat’s eye”). This phenomenon generally indicates that the subject is presently fixating at the imaging source and may be an optimal time to perform iris recognition.

Potential Consequences

Consequences emerge when vulnerabilities are associated with one or more threats. The vulnerabilities and threats identified for this project specific to the design model with an emphasis on the technical risks is listed in Table 6-2 below. The likelihood, impacts, and any mitigation plans for each consequence are addressed in the Risk Model section.

Table 6-2. Design Model Vulnerabilities, Threats, and Potential Consequences

Vulnerability	+	Threat	=	Risk
Adaptive optics have not been developed to provided sufficient quality imagery at 3m or more	+	Adaptive optics do not provide sufficient quality imagery at 3m or more.	=	**Research goals not achieved**
It is unknown if an iris can be sufficiently illuminated at a distance of 3m for use by the adaptive optics	+	Iris illumination is not sufficient at 3m or more	=	**Research goals not achieved**
Software from CMU has not been integrated with adaptive optics system	+	Software from CMU does not integrate sufficiently with adaptive optics system	=	**Research goals not achieved**

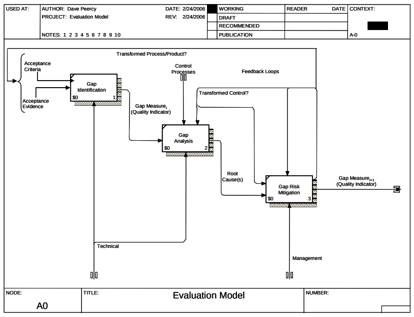
Normative References

Not specified.

References

Not specified.

6.4. Evaluation Model



The evaluation of this project will involve assessing whether or not the requirement goals were met. Gaps will be identified where the expectations of the goals were not met or could be improved upon.

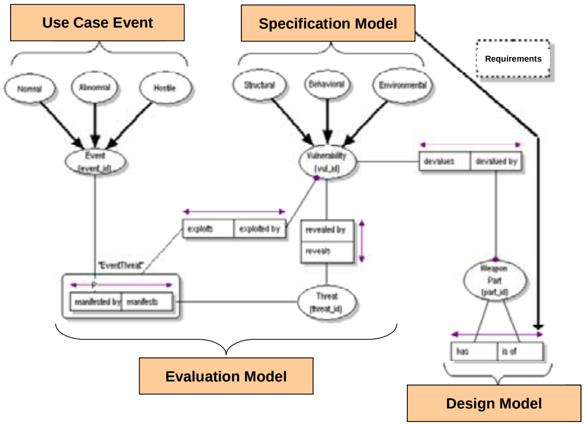
The performance of the proposed system will be validated with human-subjects data and its capabilities verified in mission-relevant scenarios. NIST Iris Recognition Grand Challenge Receiver Operator Characteristic (ROC) curves will be used as the “gold standard” for evaluating the performance of the system. The NIST numbers are for close-distance, cooperative subjects in relatively ideal conditions. These data will give us an upper bound for what is possible.

Requirement	Observation	Gap
To create an iris-recognition system for cooperative subjects at a standoff range of 3-5 meters	TBD	
To create an iris-recognition system for cooperative and uncooperative subjects at a minimum standoff range of 10 meters	TBD	

Potential Consequences

No risks in addition to those identified in the Specification and Design Models have been identified.

6.5. Risk and Maturity Models



The risks for each of the models are compiled in the table below. A probability index of high, medium or low has been assigned to each threat. This probability index can be used to determine how vital a mitigation plan might be for a project. In general, threats with a high probability should have a mitigation plan in place based on criteria of acceptance.

If no adequate mitigation plan exists, then a threat has the potential to exploit a vulnerability thus realizing the associated risk. Risks without a mitigation plan expose gaps in the surety of the project. For this Iris Recognition prototype, two gaps exist relating to project funding. While there are basic actions a Principle Investigator can perform to avoid instigating funding cuts (i.e., showing work is being performed) the decision to cut and/or reduce funds is typically beyond his/her control. Therefore, the risks relating to funding cuts exist as acceptable gaps with low probability that they will occur. Note also the three gaps related to the technical risks. While two of these risks have mitigation plans, the third one relating to the CMU software has not been addressed. It is expected that a research project will have such gaps as without risk it is unlikely to be defined as research.

The mitigation plans and all gap indicators should be reviewed periodically as the project transitions to subsequent lifecycle phases. While these mitigation plans and gaps might be acceptable for this research project within a laboratory environment, it is unlikely they will be acceptable as the Iris Recognition prototype evolves into design. The risk implications are summarized in Table 6-3. The associated Cognitive System Maturity Matrix is illustrated in Table 6-4.

Table 6-3. Case Study Risk Model Summary

Vulnerability	Threat	Like-lihood	Potential Consequence	Impact	Current Practices	Gap?	Mitigation Plan
Researchers and human subjects will be using or exposed to lab equipment	Laboratory equipment is unsafe to use	Low	**Researchers or human subjects are physically harmed**	High	ES&H Policies	No	
	Laboratory equipment is used incorrectly	Low	**Researchers or human subjects are physically harmed**	High	ES&H Policies	No	
Human subjects' iris scans will be collected	Data collected from physiological measurements are not safeguarded	High	**Human subjects' privacy might be compromised**	High	HSB review and approval	No	
Adaptive optics have not been developed to provided sufficient quality imagery at 3m or more	Adaptive optics do not provide sufficient quality imagery at 3m or more.	High	**Research goals not achieved**	High		**Yes**	Create a rapid-demonstration system in Spiral 1 to quickly test new designs using existing micro-mirrors and COTS steering mirrors Design and fabricate special-purpose micro-mirrors at MESA
Unknown if an iris can be sufficiently illuminated at a distance of 3m for use by the adaptive optics	Iris illumination is not sufficient at 3m or more	High	**Research goals not achieved**	High		**Yes**	Create a special-purpose solution in Spiral 2+

Software from CMU has not been integrated with adaptive optics system	Software from CMU does not integrate sufficiently with adaptive optics system	Medium	**Research goals not achieved**	High		**Yes**	
Project research relies on LDRD funds	LDRD funding is cut or reduced	Low	**Research goals not achieved**	Medium	Execute project communicating progress to management	**Yes**	
			Research goals not achieved	Medium	Execute project communicating progress to management		

The Cognitive System Maturity Matrix provides an at-a-glance view of the development level of the fidelity of the Iris Recognition prototype. Overall, this work is in the research stage where understanding of the concepts is the focus. There is limited attention to the consideration of an array of environments and any ethical, legal, and social implications.

Table 6-4. Case Study Cognitive System Maturity Matrix

Fidelity Level **Attribute**	**Level 0** **Low Consequence, Minimal Impact** **Scoping Studies & Research Models for Understanding**	**Level 1** **Moderate Consequence, Some Impact Preliminary Product Experimental Use**	**Level 2** **High Consequence**	**Level 3** **High Consequence, Qualified/Certified**
Psychological Representation Fidelity	Some psychological phenomena might be considered to determine how uncooperative subjects might behave in order to avoid detection.			
Physiological Representation Fidelity		Physiological measurements using calibrated test equipment.		
Environmental Representation Fidelity	Normal environment defined. No abnormal scenarios identified. No hostile scenarios identified.			
System Surety Engineering Fidelity	Limited to laboratory environment.			
System Risk Mitigation Fidelity	Limited to research project's goals.			
Ethical, Legal and Societal Implications	Discussions limited to privacy concerns regarding protection of the collection of human subjects' personal data. Misuse – no misuse scenarios have been identified. Accidental use – no accidental use scenarios have been identified. Abuse – no abusive use scenarios have been identified. Limited ethical, legal and societal implications have been identified.			

7. SUMMARY AND RECOMMENDATIONS

7.1. Summary – Research Benefits and Limitations

There is a large unknown centered on surety in the research on cognitive systems and the associated neuroscience technologies. The identification of normative references and their scientific basis is critical to the theoretical validity of such systems and technologies. The importance of such systems and technologies for use in military, national security, augmented learning systems, health-related cognitive repair and augmentation, and many other beneficial applications is apparent. However, careful consideration of the surety of such systems for their intended and unintended use is critical to their acceptance, including an understanding of ethical, legal, and societal concerns. Science without humanity[15] is indeed limiting.

The Surety Engineering Framework for Cognitive Systems provides the necessary systems engineering models to address many of the issues, concerns, and risks that have been documented in this report and the numerous references. Some aspects of this Framework and its application Process have been applied in neuroscience research associated with the brain and how it psychologically behaves and is physiologically constructed. The uniqueness of the brain being associated with our personal identity and "who we are" makes the importance of applying such an integrated systems/surety engineering and science-based approach essential.

The Surety Engineering Framework for Cognitive Systems is still fairly immature – at least in regards to use with actual cognitive system development. The concepts and most methods/techniques illustrated in this report have been extensively used in many applications – but not typically as integrated within the defined Framework and its various models. The models summarily introduced in this report and more details are known from the normative references. Only through application of the Framework and Process will value-added examples of the use for cognitive systems be derived and the Framework and Process evolved to an acceptable maturity.

7.2. Recommendations

To determine how well the Framework and Process realize the expected benefits it is necessary to apply this research to actual cognitive system projects. It would be useful to apply the Framework and Process to a variety of projects in various stages of development and implementation. The Framework and Process are easily scalable to the life cycle stage as well as to the level of complexity and project size. Specific Recommendations include:

(1) Apply the Framework and Process to one or more in-house Sandia and/or external customer cognitive system research efforts,

(2) Apply the Framework and Process to one or more military and/or national security projects addressing a cognitive system component development,

(3) Evolve the Maturity Matrix concept to include more definitive psychology and physiological cognitive system theoretical information, and

(4) Continue to research the concept of ELS engineering and determine what it means to apply the surety engineering approaches of QMU, V&V, and other such methods to this rather more subjective yet essential area.

[15] Mahatma Gandhi, one of the seven things that will destroy us.

APPENDIX A - ADDITIONAL REFERENCES

Most references in this report are contained in footnotes or in the case study sections. The following references are also relevant to this effort.

C. Althaus, “A Disciplinary Perspective on the Epistemological Status of Risk,” Risk Analysis, Volume 25, Number 3, 567, 2005.

American Medical Association, “*Principles of Medical Ethics,*” July 06, 2005.
http://www.ama-assn.org/ama/pub/category/2512.html

A. Amendola, “Recent Paradigms for Risk Informed Decision Making,” Safety Science, Volume 40, 17–30, 2001.

M. Eaton, J. Illes, “Commercializing Cognitive Neurotechnology – The Ethical Terrain,” Nature Biotechnology, Vol 25, no 4, April 2007.

G. Ersdal, T. Aven, “Risk Informed Decision-Making and Its Ethical Basis,” Reliability Engineering and System Safety, Vol 93, pp197-205, 2008.

C. Eschenbach, C. Habel, B. Smith, “Topological Foundations of Cognitive Science,” First International Summer Institute in Cognitive Science, Buffalo, New York, July 1994.

M. Farah, P. Wolpe, “Monitoring and Manipulating Brain Function: New Neuroscience Technologies and Their Ethical Implications,” Hastings Center Report 34, no. 3, pp35-45, 2004

M. Farah, “Neuroethics: the Practical and the Philosophical,” Trends in Cognitive Systems, vol 9, no 1, January 2005.

H. Greely, “*Neuroethics: The Neuroscience Revolution, Ethics, and the Law,*” Markkula Center for Applied Ethics, Santa Clara University, 2005.
http://www.scu.edu/ethics/publications/submitted/greely/neuroscience_ethics_law.html

H. Greely, “Neuroethics and National Security,” American Journal of Bioethics, May 2007.

S. Kaplan, B. Garrick, “On the Quantitative Definition of Risk,” Risk Analysis, Volume 1, Number 1, 11–27, 1981.

N. Levy, “Introducing Neuroethics,” Neuroethics, vol 1, pp 1-8, 2008.

M. McLean, “*A Framework for Thinking Ethically About Human Biotechnology,*” Markkula Center for Applied Ethics, Santa Clara University, 2005.
http://www.scu.edu/ethics/publications/submitted/mclean/biotechframework.html

J. Moreno, “*Mind Wars: Brain Research and National Defense,”* New York, NY: Dana Press.

National Academy of Sciences, “*On Being a Scientist, Responsible Conduct in Research,*” National Academy Press, Washington, DC., 1995.

National Institutes of Health, “*Regulation and Ethical Guidelines,*” Office of Human Subjects Research, December 2005.
http://www.ohsr.od.nih.gov/guidelines/guidelines.html

Research and Technology Organization, “Operator Functional State Assessment,” NATO Report, AC/323 (HFM-104)TP/48, February 2004.

D. Schmorrow, L. Reeves, A. Bolton, “21st Century Human Systems Integration: Augmented Cognition for Improved Defense Readiness,” The Technical Cooperation (TTCP) Human Systems Integration (HIS) Symposium, Australia, 2006.

W. Shaneyfelt, et. al., “Next Generation Intelligent Systems (Augmented Cognition) Grand Challenge LDRD Final Report,” SAND2006-2506, January 2006.

APPENDIX B - GLOSSARY

B.1 Acronyms

ACT-R	Adaptive Components of Thought-Rational
AMBR	Agent-Based Modeling and Behavior Representation
ART	Adaptive Resonance Theory
ASC	Advanced Simulation and Computing
ASL	Applied Science Laboratories
BE+U	Best Estimate + Uncertainty
CARION	Connectionist Learning Adaptive Rule Induction ON-line
CIT	Concealed Information Test
CMU	Carnegie Mellon University
CogSys	Cognitive System
CONOP	Concept of Operation
COSA	Cognitive System Architecture
CS&T	Cognitive Science and Technology
CSMM	Cognitive System Maturity Model
CSV	Comma Separated Value
CT	Computed Tomography (CAT Scan) ...
CVSA	Computer Voice Stress Analyzer
DAKOTA	Design Analysis Kit for Optimization and Terascale Applications
DARPA	Defense Advanced Research Projects Agency
DHMS	Distributed Human-Machine Systems
DHS	Department of Homeland Security
DIA	Defense Intelligence Agency
DNA	Deoxyribonucleic Acid
DoD	Department of Defense
DOE	Department of Energy Design of Experiments
DOH	Declaration of Helsinki
ED	Experimental Design
EEG	Electroencephalography
EI	Educing Information
ELS	Ethics, Legal, and Societal Ethics, Legal, and Sociological

EMMA	Eye Movement Memory Assessment
EPIC	Executive Process Interactive Control
ES&H	Environment Safety and Health
FDA	Food and Drug Administration
FLIR	Forward-Looking Infrared
FMEA	Failure Modes and Effects Analysis
FMECA	Failure Mode, Effects, and Criticality Analysis
fMRI	functional Magnetic Resonance Imaging
FTA	Fault Tree Analysis
FTA	Fault Tree Analysis
GKT	Guilty Knowledge Test
GMO	Genetically Modified (food)
HF	Human Factors
HF	Human Factors
HSB	Human Subjects Board
IBCA	Integrated Biologically-based Cognitive Architecture
IC	Intelligence Community
IEEE	Institute for Electrical and Electronics Engineers
IET	Integral Effects Tests
IRB	Internal Review Board
LDRD	Laboratory Directed Research and Development
LDV	Laser Doppler Vibrometry
LIDA	Learning Intelligent Distribution Agent
M&S	Modeling and Simulation
MEG	Magnetoencephalography (MEG
MIND	Mental Illness and Neurosciences Discovery
MRI	Magnetic Resonance Imaging
NARS	Non-Axiomatic Reasoning System
NE	Neuroscience Engineering
NIRS	Near Infrared Spectroscopy
NIST	National Institute of Standards and Technology
NOMAD	Neurally Organized Mobile Adaptive Device
NRC	National Research Council
NuPIC	Numenta Platform for Intelligent Computing
ONR	Office of Naval Research

ONR	Office of Naval Research
PC	Personal Computer
PCMM	Predictive Capability Maturity Model
PDA	Personal Digital Assistant
PET	Positron Emission Tomography
PM	Probabilistic Methods
QMU	Quantification of Margins and Uncertainty
R&D	Research and Development
ROC	Receiver Operator Characteristic
RVSM	Radar Vital Signs Monitor
S&T	Science and Technology
SET	Separate Effects Tests
SnePS	Semantic Network Processing System
SNL	Sandia National Laboratories
SOAR	SOAR State, Operator And Result
SRQ	System Response Quantities
TBD	To Be Determined
TMS	Transcranial Magnetic Stimulation
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNM	University of New Mexico
UQ	Uncertainty Quantification
V&U	Variation and Uncertainty
V&V	Verification and Validation
VSA	Voice Stress Analyzer

B.2 Definitions

Computational Model	A physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, or process that is implemented in a computational system.
Cognitive	The psychological and physiological processes underlying human information processing, emotion, motivation, social influence, and development.
Cognitive System	Cognitive systems are implementations of technologies that utilize as an essential component(s) one or more plausible models of human cognitive processes.
Cognitive	Products that aid a person’s cognitive functioning (comprehension,

Technology	perception, memory, problem solving and reasoning).
Ethics	The science of human duty; the body of rules of duty drawn from this science; a particular system of principles and rules concerning duty, whether true or false; rules of practice in respect to a single class of human actions; as, political or social ethics; medical ethics.
Risk	The determination of the significance of an event based on the understanding of: (1) What scenario can happen under which the event would occur; (2) What is the likelihood that conditions for the event will occur; and (3) What are the consequences if the event were to occur. The determination may depend on who decides the significance of the event information – risk agent. Risk in a system is directly associated with the potential to lose an asset of the system: the existence of a system vulnerability to adequately protect specified assets under a credible scenario and the potential existence of a threat that could take advantage of the vulnerability. Risks can be categorized in many different ways – such as illustrated in the categories below.
	Subjective risk: the mental state of an individual who experiences uncertainty or doubt or worry as to the outcome of a given event.
	Objective risk: the variation that occurs when actual losses differ from expected losses.
	Real risk: the combination of probability and negative consequence that exists in the real world.
	Observed risk: the measurement of that combination obtained by constructing a model of the real world.
	Perceived risk: the rough estimate of real risk made by an untrained member of the general public.
Risk Dimensions	The following summary of risk dimensions is derived from [ALTHAUS-2005].
	(1) linguistic and conceptual: concerned with the semantic variances in meaning of the term risk and its variability in societal use.
	(2) historical and narrative: concerned with the historical evolution of the use of risk as a phenomenon in its own right in particular, the discoveries in mathematics, economics, and psychology that enabled risk to be better understood and measured.
	(3) mathematical and logical: concerned with the definition and development of risk in mathematical and logical terms; risk is a function of probability, which can be derived from statistics and that can be modeled with game theory.
	(4) scientific and measurable: concerned with specific application to areas such as science disciplines to understand and define risk as an objective reality that can be measured, controlled, and managed. Develops notion of prediction of hazards and judgment as to what is “acceptable” risk.

	(5) economic and decisional: concerned with the application of risk methods to economic applications to provide a basis for making decisions that affect wealth. The general concept of risk in economics is a mix of challenge and security. The predominant focus is that of the risk-reward paradigm that represents the voluntary and incentive perspective of risk.
	(6) psychological and cognitive: concerned with the subjective nature of risk vs the objective scientific view of risk; risk perception vs risk action on a cognitive level. Concerned with determining why there is disparity between expert and lay risk perception; what makes people risk-averse, risk-indifferent, or risk-takers; explores the significance of trust, blame, vulnerability, defense mechanisms, and other aspects of motivation and cognition that characterize risk behavior.
	(7) anthropological and cultural: concerned with why people worry about different risks and whether risk is actually increasing. The key question raised by anthropology is why does risk analysis not take into account cultural issues when discussing risk? As soon as culture is introduced, risk becomes politicized; a cultural perspective wrenches risk from its scientific and mathematical foundations by positing risk to be a choice word for danger.
	(8) sociological and societal: concerned with the rippling undercurrent of risk as a form of humanism: does humanity have the capacity to determine its future, does it trust itself, or will impending technical catastrophes overrun the human spirit? Risk and society are fundamentally and inextricably intertwined. Risk can be understood as a societal phenomenon. It explains, shapes, delineates, and defines society and vice versa. Only with risk can we understand society and only with society can we understand risk.
	(9) artistic and emotional: concerned with risk as the possibility of isolation or the possibility of connection; encapsulates both the danger and venturesome meanings of risk as an emotion-based description.
	philosophical and phenomenological: concerned with establishing the ontological foundation of risk. Also concerned with epistemology, where debates on risk rage over the question of experts and the relative position of ignorance and knowledge vis-a-vis risk: who can we trust, who are the experts, how should expert knowledge be applied, how does ignorance and knowledge impact on risk decisions, how does truth and error pertain to risk?

	(11) legal and judicial: concerned with the assumption that damage or harm can occur even when the defendant is not at fault and it is more the “exposure to risk” that is offensive for the purposes of guilt and injustice. Encompasses and broadens the interpretation of a defendant is at fault if it can be shown that intention and negligence were present, with negligence historically based on the notion of the average, reasonable, competent person. This leads to a broad legal and judicial interpretation of liability – both corporate and individual.
	(12) theological: concerned with the moral dimension of risk and its effect on the human spirit and specific religious rules; analysis of the treatment of entrepreneurship in Religions, concluding on the vocational aspects of entrepreneurship and the positive moral dimension to risk-taking behavior. Risk in theology is an act of faith. In applying calculations to the unknown, for example, mathematics establishes one definitional perspective on risk that renders risk a calculable phenomenon. In applying revelation to the unknown, religion establishes another perspective on risk that views it in light of faith.
Surety Technology	Methods and techniques that are applied from the disciplines of safety, reliability, security/use control, human factors, surveillance, and quality. There are other methods of providing surety that may focus on organizational structure, societal checks and balances, and other such non-technical approaches.
Surety Framework Definitions	
Environment Scenario	Sequence of activities through which the product (system, subsystem, component) is intended to satisfy its specifications in accordance with how it has been designed Instance: intended and/or unintended scenarios for cognitive system use.
Normative References	Standards, historical evidence/lessons learned, and/or expert opinion that represents process and/or product best practices for any of the other elements of the framework Instance: safety first principles, ethical principles, psychological theories and standard models, physiological structures and standard models, state-of-the-art tools (e.g., fMRI) that might be used as part of or for evaluation of a cognitive system.
Specification Model	Generic requirements (behavioral, structural, environmental) that address the class of products/processes (in our case, cognitive systems) within the scope of the quality framework Instance: requirements of a specified cognitive system product/process
Design Model	Generic architecture (physical/functional) that describes the class of products (in our case, cognitive systems) within the scope of the framework. Instance: design and processes used for a specified cognitive system product; use of tools (e.g., episodic memory model) as part of design.

Evaluation Model	Generic processes and methods that might be used to obtain measures of how well the requirements of the specification model are met by the architecture of the design model. Instance: specific verification and validation experiments with QMU analysis to obtain measures of how well the requirements of a cognitive system product/process are met; use of tools (e.g., Design of Experiment, Statistical Process Control, vibration/shock/temperature testing processes and equipment, fMRI) as part of evaluation methods.
Risk Model	Generic gap analysis processes and methods that might be used in a time/phase-dependent approach to determine the implications of the gap measures obtained by the Evaluation Model. Risk-informed approach to managing vulnerabilities to a targeted risk. Instance: risk-based analysis (potential threat, impact of threat/vulnerability occurrence, and likelihood of occurrence) of a cognitive system's lack of a safety theme implementation, or lack of an authentication mechanism to prevent exploitation of cognitive system privacy information, or a process gap in the conduct of external peer review of the ELS concerns for a new neuroscience technology; representation of the gap/risk indicators in a risk-informed prioritization based on the risk aversion threshold and perhaps the life cycle stage of the cognitive system.
Maturity Model	Level of maturity of the cognitive system model as it relates to potential application use. Plausibility characteristics of the cognitive system (psychological, physiological, environmental conditions) as they relates to potential application use. Instance: research, early prototype development, full scale development & production, high consequence (regulatory) qualification application represent stages for which maturity/plausibility criteria might apply; measure of how well an episodic memory model actually represents human cognition.

APPENDIX C - EXISTING COGNITIVE SYSTEM ARCHITECTURES AND EMERGING TECHNOLOGIES

C.1 Existing Cognitive System Architectures

Some of the existing cognitive system architectures that provide instances of cognitive system design models (or at least representations of attempts at implementing some aspect of cognitive systems design models) are illustrated and a simple categorization[16] provided in the Table C-1. Memory (e.g., episodic, special, semantic) and Learning (e.g., contextual knowledge, emotional process, comparison processes) are two primary areas that distinguish architectures within this simple classification scheme.

- **Emergent:** architectures that use low-level activation signals flowing through a network consisting of numerous processing units; a bottom-up process relying on the emergent self-organizing and associated properties; memory representation has global/local elements and learning representation is associative/competitive.
- **Symbolic:** architectures that use symbols as the key means to support information processing; memory representation is rule/graph-based and learning representation is inductive/analytical.
- **Hybrid:** architectures that results from combining the symbolic and emergent paradigms in one way or another; memory representation tends to be a combination of localist-distributed and/or symbolic-connectionist; learning representation tends to be a combination of bottom-up and/or top-down.

The Surety Engineering Framework Design Model is essentially a more detailed Hybrid approach, particularly the architectural extensions as indicated in Appendix F.2.

Table C-1. Examples of Existing Cognitive System Architectures

Instance	Description	Type	Ref
Cortronics	An emergent architecture that models the biological functions of the cerebral cortex and thalamus systems (jointly termed thalamocortex) in the human brain. Its memory organization consists of modular feature attractor circuits called lexicons. Each lexicon comprises further a localist cortical patch, a localist thalamic patch, and the reciprocal connections linking them. Knowledge takes the form of parallel, indirect unidirectional links between the neurons representing one symbol in a lexicon and those describing a symbol in another lexicon. Each such knowledge link is termed an item of knowledge, and the collection of all these links is called a knowledge base. The union of cortical patches of all lexicons	Emergent	17

[16] W. Duch, R. Oentaryo, M Pasquier, "Cognitive Architectures: where do we go from here?," Department of Informatics, Nicolaus Copernicus University, Grudziadzka 5, 87100 Torun, Poland.

[17] R. Hecht-Nielsen, Confabulation Theory: The Mechanism of Thought. Springer 2007.

	constitutes in turn the entire cortex, while that of the thalamic patches of all lexicons forms only a portion of thalamus. A competitive activation of symbols of lexicons, called confabulation, is used for learning and information retrieval. This is quite new architecture and it is not yet clear how it can be extended to create generalized intelligence, as confabulation is not sufficient for reasoning with complex knowledge.		
IBCA	*Integrated Biologically-based Cognitive Architecture* A large-scale emergent architecture that epitomizes the automatic and distributed notions of information processing in the brain. The role of three regions in the brain is emphasized: posterior cortex (PC), frontal cortex (FC), and hippocampus (HC). The PC module assumes an overlapping, distributed localist organization that focuses on sensory-motor as well as multi-modal, hierarchical processing. The FC module employs a non-overlapping, recurrent localist organization in which working memory units are isolated from one another and contribute combinatorially (with separate active units representing different features). The HC module utilizes a sparse, conjunctive globalist organization where all units contribute interactively (not combinatorially) to a given representation. Representation of emotions for motivation and setting goals, as well as motor coordination and timing, is still missing.	Emergent	[18]
NOMAD	*Neurally Organized Mobile Adaptive Device* Nomads, also known as Darwin automata, demonstrate the principles of emergent architectures for pattern recognition task in real time. They use many sensors for vision, range finders to provide a sense of proximity, prioproceptive sense of head direction and self-movement, artificial whiskers for texture sensing, artificial taste (conductivity) sensors. NOMAD is controlled by a large simulated nervous system running on a cluster of powerful computers. the emergence of higher-level cognition does not seem likely in this architecture.	Emergent	[19]
NuPIC	*Numenta Platform for Intelligent Computing* Based on the Hierarchical Temporal Memory (HTM) technology, which is modeled on the putative algorithm used by neocortex. Network nodes are organized in a hierarchical way, with each node implementing learning and memory functions. Hierarchical organization is motivated by the growing size of	Emergent	[20]

[18] R. O'Reilly, T. Braver, J. Cohen, "A Biologically-Based Computational Model of Working Memory," A. Miyake & P. Shah (Eds.), Models of Working Memory. Cambridge University Press, pp. 375-411, 1999.

[19] .M. Edelman, "Neural Darwinism: Selection and Reentrant Signaling in Higher Brain Function," Neuron 10, pp115-125, 1993.

[20] J. Hawkins, S. Blakeslee, On intelligence: How a New Understanding of the Brain will Lead to the Creation of Truly Intelligent Machines. Times Books 2004.

	cortical receptive fields in the information streams that connect primary sensory cortices with secondary and higher-level association areas. This feature is also present in the IBCA architecture, where specific connectivity between different layers leads to growing and invariant object representation. The architecture has not yet been tested in larger applications.		
EPIC	*Executive Process Interactive Control* Architecture for building computational models include many aspects of human performance. Aims at capturing human perceptual models of human-computer interaction. System is controlled by production rules for cognitive processor and a set of perceptual (visual, auditory, tactile) and motor processors operating on symbolically coded features rather than raw sensory data. EPIC-SOAR combination has been applied to air traffic control simulation.	Symbolic	[21]
ICARUS	*ICARUS Project* Integrated cognitive architecture for physical agents, with knowledge specified in the form of reactive skills, each denoting goal-relevant reactions to a class of problems. Concepts are matched to percepts in a bottom-up way and goals are matched to skills in a top-down way. Conceptual memory contains knowledge about general classes of objects and their relationships, while skill memory stores knowledge about the ways of doing things. Each comprises a long-term memory (LTM) and a short-term memory (STM). The acquisition of knowledge in ICARUS is achieved through hierarchical, incremental reinforcement learning, propagating reward values backward through time.	Symbolic	[22]
NARS	*Non-Axiomatic Reasoning System* A reasoning system based on a language for knowledge representation, an experience-grounded semantics of the language, a set of inference rules, a memory structure, and a control mechanism, carrying out various high level cognitive tasks as different aspects of the same underlying process.	Symbolic	[23]
SnePS	*Semantic Network Processing System* Is a logic, frame and network-based knowledge representation, reasoning, and acting system that went through over three decades of development. It stores knowledge and beliefs of some agent in form of assertions (propositions) about various entities.	Symbolic	[24]

[21] D. Meyer, D. Keiras, "A Computational Theory of Executive Cognitive Processes and Multiple-task Performance: Part 1. basic Mechanisms." Psychological Review, 10(1) pp 3-65, 1997.

[22] P. Langley, "An adaptive architecture for physical agents," Proceedings 2005 IEEE/WIC/ACM International Conference on Intelligent Agent Technology. Compiegne, France: IEEE Computer Society Press, pp. 18-25, 2005.

[23] P. Wang, "Rigid Flexibility. The Logic of Intelligence," Springer 2006.

[24] S. Shapiro, W. Rapaport, M. Kandefer, F. Johnson, A. Goldfain, "Metacognition in SNePS," AI Magazine 28 pp 17-31, 2007.

	Each knowledge representation has its own inference scheme, logic formula, frame slots and network path.. When a belief revision system discovers contradiction some hypotheses that led to the contradiction have to be unasserted by the user and the system removes all those assertions that depend on it. Has been used for common-sense reasoning, natural language understanding and generation, contextual vocabulary acquisition, control of simulated cognitive agent that is able to chat with the users, question/answer system and other applications. Interesting inferences have been demonstrated, but the program has not yet been used in a large-scale real application.		
SOAR	*State, Operator And Result* Classic expert rule-based cognitive architecture designed to model general intelligence. Based on theoretical framework (normative reference) of knowledge-based systems arranged in terms of operators that act in the problem space – the set of states that represent the immediate task at hand. The primary explanation-based learning technique for formulating rules and macro-operations from problem solving traces. In recent years many extensions of the architecture have been proposed: reinforcement learning to adjust the preference values for operators, episodic learning to retain history of system evolution, semantic learning to describe more abstract, declarative knowledge, visual imagery, emotions, moods and feelings used to speed up reinforcement learning and direct reasoning. The architecture has demonstrated a variety of high-level cognitive functions, processing large and complex rule sets in planning, problem solving and natural language comprehension (NL-SOAR) in real-time distributed environments. At present the architecture has not yet integrated all these extensions. A few additional ones, like memory decay/forgetting, attention and information selection, learning hierarchical representations, or handling uncertainty and imprecision, will also be useful. The design of the perceptual-motor systems is fairly unrealistic, requiring users to define their own input and output functions for a given domain.	Symbolic	[25]
4CAPS	Has plausible neural implementation and is designed for complex tasks, such as language comprehension, problem solving or spatial reasoning. A unique feature is the ability to compare the activity of the different architecture modules with functional neuroimaging measures of brain's activity. It has been used to model human behavioral data (response times and error rates) for analogical problem solving, human-computer	Hybrid	[26]

[25] A. Newell, Unified Theories of Cognition," Harvard University Press, 1990.

[26] M. Just, S. Varma, "The Organization of Thinking: What Functional Brain Imaging Reveals About the Neuroarchitecture of Complex Cognition," Cognitive, Affective, and Behavioral Neuroscience, 7, pp 153-191, 2007.

	interaction, problem solving, discourse comprehension and other complex tasks solved by normal and mentally impaired people. Its first operating principle, “Thinking is the product of the concurrent activity of multiple centers that collaborate in a large scale cortical network”, leads to the architecture based on a number of centers (corresponding to particular brain areas) that have different processing styles. Contains many interesting ideas but it is not aimed at achieving intelligent behavior - rather tries to model human performance.		
ACT-R	*Adaptive Components of Thought-Rational* A hybrid cognitive architecture and theoretical framework for emulating and understanding human cognition. Aims at building a system that can performs the full range of human cognitive tasks and describe in detail the mechanisms underlying perception, thinking, and action. The central components comprise a set of perceptual-motor modules, memory modules, buffers, and a pattern matcher. utilizes a top-down learning approach to adapt to the structure of the environment. This architecture may be partially mapped on the brain structures. It has been applied in a large number of psychological studies, and in intelligent tutoring systems, but ambitious applications to problem solving and reasoning are still missing.	Hybrid	[27] [28]
AMBR	*Agent-Based Modeling and Behavior Representation* A model of human reasoning that unifies analogy, deduction, and generalization, including a model of episodic memory; a model of human judgment; a model of perception, analysis of interactions between analogy, memory, and perception; understanding the role of context and priming effects for the dynamics of cognitive processes. This is certainly a very interesting architecture that is capable of explaining many cognitive phenomena. A recent project resulted in quantitative data comparing the performance of humans and cognitive architectures in a simplified air traffic controller environment. It is not clear how well it will scale up to real problems requiring complex reasoning, as nothing in this area has yet been demonstrated.	Hybrid	[29]
CLARION	*Connectionist Learning Adaptive Rule Induction ON-line* Incorporates a distinction between explicit (symbolic) and implicit (sub-symbolic) processes and captures the interactions	Hybrid	[30]

[27] J. Anderson, C. Lebiere, “The Newell test for a Theory of Cognition,” Behavioral and Brain Science 26, pp 587-637, 2003.

[28] K. Gluck and R. Pew (editors), Modeling Human Behavior with Integrated Cognitive Architectures, Lawrence Erlbaum Associates, New Jersey, 2005.

[29] K. Gluck and R. Pew (editors), ibid.

[30] R. Sun, X. Zhang, “Top-Down Versus Bottom-Up Learning in Cognitive Skill Acquisition,” Cognitive Systems Research 5, pp 63-89, 2004.

	between the two. Architecture contains four memory modules, each comprising a dual explicit-implicit representation: action-centered subsystem , non-action-centered subsystem, motivational subsystem, and metacognitive subsystem. Each of these modules adopts a localist-distributed representation, where the localist section encodes the explicit knowledge and the distributed section the implicit knowledge. Employs different learning methods for each level of knowledge. A lot of psychological data has been simulated with this architecture, but also a complex sequential decision-making for a minefield navigation task. There is no doubt that this architecture may explain many features of mind, however, it remains to be seen how much competence it will achieve in understanding language, vision, and common sense reasoning based on perceptions.		
COSA	*Cognitive System Architecture* A generic framework proposing a unified architecture for cognitive systems. A new engineering approach for cognitive systems, implemented by the architecture, which may be a crucial step forward to achieve a wide-spread application of cognitive systems. The approach is based on a new concept of generating cognitive behavior, called the cognitive process (CP). The CP can be regarded as a model of the human information processing loop whose behavior is solely driven by ‘‘a-priori knowledge’’. The main features are the implementation of the CP as its kernel and the separation of architecture from application leading to reduced development time and increased knowledge reuse. Additionally, separating the knowledge modeling process from behavior generation enables the knowledge designer to use the knowledge representation that is best suited the modeling problem. A first application based on implements an autonomous unmanned air vehicle accomplishing a military reconnaissance mission.	Hybrid	[31]
DUAL	This architecture as been inspired by Minsky’s “Society of Mind” theory of cognition. It is a hybrid, multi-agent general-purpose architecture supporting dynamic emergent computation, with a unified description of mental representation, memory structures, and processing mechanisms carried out by small interacting micro-agents. As a result of lack of central control the system is constantly changing, depending on the environment. Agents interact forming larger complexes, coalitions and formations, some of which may be reified. Such	Hybrid	[32]

[31] H. Putzer, R. Onken, “COSA – A Generic Cognitive System Architecture Based on a Cognitive Model of Human Behavior,” Cognition, Technology & Work, 5(2), pp 140-151, 2003.

[32] A. Nestor, B. Kokinov, “Towards Active Vision in the DUAL Cognitive Architecture,” International Journal on Information Theories and Applications, 11, pp 9-15, 2004.

	models may be evaluated at different levels of granularity, the microlevel of micro-agents, the mesolevel of emergent and dynamic coalitions, and the macrolevel of the whole system and models, where psychological interpretations may be used to describe model properties. Micro-frames are used for symbolic representation of facts, while relevance or activation level of these facts in a particular context is represented by network connections with spreading activation that changes node accessibility. This architecture has been used in a number of projects.		
LIDA	*Learning Intelligent Distribution Agent* A conceptual and computational framework for intelligent, autonomous, "conscious" software agent that implements some ideas of the global workspace theory. The architecture is built upon a bit older IDA framework, which was initially designed to automate the whole set of tasks of a human personnel agent who assigns sailors to new tours of duty. Employs a partly symbolic and partly connectionist memory organization, with all symbols being grounded in the physical world. Has distinct modules for perception, working memory, emotions, semantic memory, episodic memory, action selection, expectation and automatization (learning procedural tasks from experience), constraint satisfaction, deliberation, negotiation, problem solving, metacognition, and conscious-like behavior.	Hybrid	[33]
Novamente AI Engine	Based on system-theoretic ideas regarding complex mental dynamics and associated emergent patterns, inspired by the psynet model and more general "patternist philosophy of mind". Similarly as in the "society of minds" and the global workspace, self-organizing and goal-oriented interactions between patterns are responsible for mental states. Emergent properties of network activations should lead to hierarchical and relational (heterarchical) pattern organization. Probabilistic Term Logic, and the Bayesian Optimization Algorithm(s) are used for flexible inference. Actions, perceptions, and internal states are represented by tree-like structures. This is still an experimental architecture that is being developed, seems to be in a fluid state, and its scaling properties are not yet known.	Hybrid	[34] [35]
Polyscheme	Integrates multiple methods of representation, reasoning and inference schemes in problem solving. Each "specialist" models	Hybrid	[36]

[33] S. Franklin, "The LIDA Architecture: Adding New Modes of Learning to an Intelligent, Autonomous, Software Agent," In Proceedings. of the International Conference on Integrated Design and Process Technology, San Diego, CA: Society for Design and Process Science, 2006.

[34] B. Goertzel, From Complexity to Creativity, New York, NY: Plenum Press, 1997.

[35] B. Goertzel, The Hidden Pattern, BrownWalker Press, 2006.

[36] N. Cassimatis, Adaptive Algorithmic Hybrids for Human-Level Artificial Intelligence, Advances in Artificial General Intelligence. IOS Press. Eds. B. Goertzel and P. Wang, 2007.

	a different aspect of the world using specific representation and inference techniques, interacting with other specialists and learning from them. Scripts, frames, logical propositions, neural networks and constraint graphs can be used to represent knowledge. A reflective specialist guides the attention of the whole system, providing various focus schemes that implement inferences via script matching, backtracking search, reason maintenance, stochastic simulation and counterfactual reasoning. May be used both in abstract reasoning and also in common sense physical reasoning in robots. It has been used to model infant reasoning including object identity, events, causality, spatial relations. This meta-learning approach combining different approaches to problem solving is certainly an important step towards AGI and common sense reasoning.		
Shruti	Biologically-inspired model of human reflexive inference, represents in connectionist architecture relations, types, entities and causal rules using focal-clusters. These clusters encode universal/existential quantification, degree of belief, and the query status. The synchronous firing of nodes represents dynamic binding, allowing for representations of quite complex knowledge and inferences. This architecture may have great potential, but after rather long time of development it has not yet found any serious applications to problem solving or language understanding.	Hybrid	[37]

Some conclusions can be derived from the various architecture reference material:

(1) architectures are really instances of the Surety Engineering Framework Design Model

(2) architectures are based on specific normative reference theories, principles, or organizational methods that reflect one or more aspects of the cognitive capabilities represented in psychology and/or physiology of the brain

(3) although the architectures reference specific theories/principles as the normative basis, there seems to be a lack of abstraction for the theories/principles that might provide broader coverage across all the architectures; the Langley reference[38] provides some very interesting challenges and suggestions in defining this normative reference abstraction

(4) architectures have reasonably specific focus areas of interest; very few take a more global view of the cognitive system concepts – primarily because of the complexity of the problem of representing human cognition

(5) architectures have quite a lot of commonality in concepts although the implementing mechanisms may differ

[37] L. Shastri, V. Ajjanagadde, “From Simple Associations to Systematic Reasoning: A Connectionist Encoding of Rules, Variables, and Dynamic Bindings Using Temporal Synchrony,” Behavioral & Brain Sciences, 16(3), pp 417-494, 1993.

[38] P. Langley, J. Laird, S. Rogers, “Cognitive Architectures: Research Issues and Challenges,” pre-publication copy, submitted to Elsevier, September 23, 2008.

(6) considerations of safety occasionally are mentioned, security/privacy concerns are not generally addressed, and the issues surrounding ethics, legal, and societal concerns of the architectures' applications does not seem to be a serious discussion point

(7) although environment scenarios (intended use/unintended use) are sometimes described in terms of specific application use – a more general consideration of the issues surrounding more comprehensive environmental concerns does not seem to be a focus

(8) architecture implementations generally lack a systems engineering approach to the requirements specification, design definition, and implementation verification and validation; however, almost all acknowledge the lack of validation evidence of capabilities and the inability to adequately conduct such necessary experiments

(9) there is no common systematic approach to understanding the validity of the normative reference theories, the application validity of how well the architecture requirement specifications cover the theory, nor how well the architecture design actually implements the requirement specification through application of systems/surety engineering methods and techniques; in short, there is no evidence of systematic application of systems engineering principles and methods

The Surety Engineering Framework for Cognitive Systems provides a very reasonable approach to encompassing the existing architectures and research work. In addition the Framework addresses the lack of a systematic systems/surety engineering approach to determining the validation of the normative theories, maturity of the cognitive system architecture, fidelity of the architecture's implementation, and the associated areas of risk associated with the technical fidelity and how well the implementation has addressed potential ELS issues. The following quote[39] illustrates the need for such an engineering framework for application to cognitive systems research, implementation, and productization – although not only for software implementations, but of the whole systems (hardware, software, developers, users, suppliers) approach.

"Conventional automation and similar systems lack the ability of cooperation and cognition, leading to serious deficiencies when acting in complex environments, especially in the context of human-computer interaction. Cognitive systems based on cognitive automation can overcome these deficiencies. Designing such artificial cognitive systems can be considered a very complex software development process. Although a number of developments of artificial cognitive systems have already demonstrated great functional potentials in field tests, the engineering approach of this kind of software is still a candidate for further improvement. Therefore, wide-spread application of cognitive systems has not been achieved yet."

The Appendix C - Section C.2 summarizes more of these concerns and how the Surety Engineering Framework for Cognitive Systems can be applied to emerging cognitive neuroscience and related technologies. In particular, relevance of the integration and engineering of surety and ELS concerns is addressed.

[39] H. Putzer, R. Onken, "COSA – A Generic Cognitive System Architecture Based on a Cognitive Model of Human Behavior," Cognition, Technology & Work, 5(2), pp 140-151, 2003.

C.2 Emerging Cognitive Neuroscience and Related Technologies

As part of the National Academies of Science mission to educate the world on issues of science, engineering, and health a report[40] on Emerging Cognitive Neuroscience and Related Technologies has been compiled. This very informative report was created by the Committee on Emergent Neurophysiological and Cognitive/Neural Science Research in the Next Two Decades as tasked by the Technology Warning Division of the Defense Intelligence Agency's (DIA's) Defense Warning Office to identify areas of cognitive neuroscience and related technologies that will develop over the next two decades and that could have military applications that might also be of interest to the Intelligence Community (IC). Specifically, the DIA asked the National Research Council (NRC) to perform the followings tasks:

(1) Review the current state of today's work in neurophysiology and cognitive/neural science, select the manners in which this work could be of interest to national security professionals, and identify trends for future warfighting applications that may warrant continued analysis and tracking by the intelligence community,

(2) Use the technology warning methodology developed in the 2005 National Research Council report[41] to assess the health, rate of development, and degree of innovation in the neurophysiology and cognitive/neural science research areas of interest, and

(3) Amplify the technology warning methodology to illustrate the ways in which neurophysiological and cognitive/neural research conducted in selected countries may affect committee assessments.

The label "cognitive" in the report is used in a broad sense. It is reflective of the "cognitive sciences" in general to refer to psychological and physiological processes underlying human information processing, emotion, motivation, social influence, and development. Hence, it includes contributions from all directly related cognate disciplines including behavioral and social science disciplines as well as contributing disciplines such as philosophy, mathematics, computer science, and linguistics. The label "neuroscience" is also used in a broad sense and includes the central nervous system (e.g., brain), and the somatic, autonomic, and euroendocrine processes.

Report Bottom Line: "Cognitive neuroscience and its related technologies are advancing rapidly, but the IC has only a small number of intelligence analysts with the scientific competence needed to fully grasp the significance of the advances. Not only is the pace of progress swift and interest in research high around the world, but the advances are also spreading to new areas of research, including computational biology and distributed human–machine systems with potential for military and intelligence applications. Cognitive neuroscience and neurotechnology comprise a multifaceted discipline that is flourishing on many fronts. Important research is taking place in detection of deception, neuropsychopharmacology, functional neuroimaging, computational biology, and distributed human-machine systems, among other areas. Accompanying this research are the ethical and cultural implications and considerations

[40] National Research Council, "Emerging Cognitive Neuroscience and Related Technologies," prepublication copy, Committee on Military and Intelligence Methodology for Emergent Neurophysiological and Cognitive/Neural Science Research in the Next Two Decades, Division on Engineering and Physical Sciences, National Academies Press, Washington, DC, 2008.

[41] National Research Council, "Avoiding Surprise in an Era of Global Technology Advances," National Academies Press, Washington, D.C., 2005

that will continue to emerge and will require serious thought and actions. The IC also confronts massive amounts of pseudoscientific information and journalistic oversimplification related to cognitive neuroscience."

Framework Relevance: Due to the complexity and extensiveness of the cognitive neuroscience and related technologies research, a systematic approach is needed to:

(1) separate out pseudoscientific and over-simplified information (e.g., non-evidence-based research and research whose evidence does not support its claims),

(2) integrate ethical and cultural implications and considerations.

(3) identify the maturity of a multitude of emerging technologies from around the world,

(4) address serious military and national security challenges, and

(5) augment/improve the technical capabilities of the technology warning methodology.

To address all these new technology concerns it is necessary to adopt a common framework that can be scaled to a large variety of applications, incorporates a systems engineering discipline, applies known as well as innovative methods and techniques for verification and validation of the technical requirements for reliability, safety, and security, and incorporates a disciplined (engineering) approach to elicit and mitigate risks due to ELS concerns. The Surety Engineering Framework for Cognitive Systems should be able to address these challenges as the approach is applied to specific applications and the Framework is improved and evolved.

Specific Report Findings and Framework Application Potential

The following key findings and their relationship to the Surety Engineering Framework for Cognitive Systems will be briefly described. Only those findings that are specifically pertinent to the application of the Framework will be discussed.

Key Finding 5-5: The recommendations in this report to improve technology warning cognitive neuroscience and related technologies are unlikely to succeed unless the following issues are addressed:

(1) Emphasizing science and technology as a priority for intelligence collection and analysis.

(2) Appointing and retaining accomplished IC professionals with advanced scientific and technical training to aid in the development of S&T collection strategies.

(3) Increasing external collaboration by the IC with the academic community.

Framework Relevance: The foundation of the Framework's approach is science-based and risk-informed with the specified models to accomplish this. The Framework provides a common approach to measure the progress and maturity of cognitive neuroscience and related technologies to improve the technology warning process.

Key Finding 2-2: The committee recognizes the IC's strong interest in improving its ability to detect deception. Consistent with the 2003 NRC study on polygraphs and lie detection[42], the committee uniformly agreed that, to date, insufficient, high-quality research has been conducted

[42] National Research Council (NRC), Polygraph and Lie Detection, National Academies Press, Washington, D.C, 2003.

to provide empirical support for the use of any single neurophysiological technology, including functional neuroimaging, to detect deception.

Accompanying this finding was a committee recommendation:

Key Recommendation 2-1: The committee recommends further research on multimodal methodological approaches for detecting and measuring neurophysiological indicators of psychological states and intentions. This research should combine multiple measures and assessment technologies, such as imaging techniques and the recording of electrophysiological, biochemical, and pharmacological responses. Resources invested in further cognitive neuroscience research should support programs of research based on scientific principles and that avoid the inferential biases inherent in previous research in polygraphy.

The committee had a specifically pertinent statement concerning ethics:

> *"Importantly, human institutional review board standards require, at minimum, that individuals not be put at any greater risk than they would be in their normal everyday lives. The committee believes certain situations would allow such testing under "normal risk" situations; though the committee strongly endorses the necessity of realistic, but ethical, research in this area, it does not specify the nature of that research in this report."*

Framework Relevance: The conclusion of the committee that "the use of any single neurophysiological technology" was inadequate based on the research emphasizes several fundamental principles of surety engineering. First, in order to achieve higher reliable results it is usually necessary to have multiple mechanisms that would have to fail in order for overall system failure to occur. Hence, the use of multiple techniques (perhaps existing ones, perhaps new technologies) that have been integrated into a well-defined and engineered process would be one normative standard way to improve the existing research results. Second, another principle of safety is that there is no single point of failure, and that the multiple points that would have to fail are to be independent and to some extent "layered". In the context of detection deception it would mean there are multiple mechanisms (perhaps from simple to complex) that would be applied to provide higher confidence that deception is detected when present and non-detected when it is not present. In order to reduce behavioral and/or structural failures within potentially multiple environments, the associated implementations of the neurophysiological technologies must be isolated from potential sources of failure and incompatible (e.g., electrically) with any potential sources of failure. These are examples of scientific principles upon which the committee recommends further research depend. Further discussion of detection deception within the broader context of Educing Information is briefly discussed in Appendix F - Section F.1.

Key Finding 2-5. Functional neuroimaging is progressing rapidly and is likely to produce important findings over the next two decades. For the intelligence community and the Department of Defense, two areas where such progress could be of great interest are enhancing cognition and facilitating training. Additional research is still needed on states of emotion; motivation; psychopathology; language; imaging processing for measuring workload performance; and the differences between Western and non-Western cultures.

Framework Relevance: The research summary provided in Appendix F - Section F.2 provides an example how the Framework can be applied to an expanded and higher fidelity Design Model where such additional research (not specifically for fMRIs but in related cognitive areas) is being performed – including investigation of non-Western cultures. Some key considerations would be to validate existing fMRI performance – not just for the imaging aspects but for the specific application domains of interest such as detection deception and more general understanding of cognitive processing. Unintended applications and associated ELS issues (e.g., see Appendix D - Section D.2) such as proposed for use with the Framework must be part of the solution. This is supported by the following committee statement:

> *"Functional neuroimaging technologies are commonplace in research and clinical environments and are affecting defense policy. Their continued development and refinement are likely to lead to applications that go well beyond those envisioned* by current cognitive neuroscience research and clinical medicine."

This application area may well be critically important to the military and national security – specifically the Intelligence Community - but not without the surety engineering required to provide the adequate confidence that potential risks in use have been mitigated. Some of the potential applications include:

(1) providing insight into intelligence from captured military combatants,

(2) enhancing military training techniques,

(3) enhancing cognition and memory of enemy soldiers and intelligence operatives,

(4) screening terrorism suspects at checkpoints or ports of entry, and

(5) improving the effectiveness human–machine interfaces in such applications as remotely piloted vehicles and prosthetics.

Key Finding 3-6: As high-performance computing becomes less expensive and more available, a country could become a world leader in cognitive neurosciences through sustained investment in the nurture of local talent and the construction of required infrastructure. Keys to allowing breakthroughs will be the development of software-based models and algorithms where much of the world is now on par with or ahead of the United States. Given the proliferation of highly skilled software researchers around the world and the relatively low cost of establishing and sustaining the necessary organizational infrastructure in many other countries, the United States cannot expect to easily maintain its technical superiority.

The committee provided an accompanying recommendation to develop the capability to monitor international progress and investments in computational neuroscience.

Framework Relevance: The use of high-performance computing is an integrated part of the Sandia neuroscience research capability. More important from the Framework perspective is the integration of the Modeling & Simulation (M&S) methodologies for Verification and Validation, Quantification of Margins and Uncertainties, rigorous qualification of software-based models and algorithms using the computational engineering layering of theory, science (e.g., physics, biological, chemical science-based models), mathematics, numerical methods, and software implementation within a comprehensive V&V approach[43]. See Appendix E - all sections for examples of surety engineering methods and techniques.

Key Finding 3-7: Unlike in the domain of cognitive neurophysiological research, where the topics are constrained by certain aspects of human physiology and brain functioning, progress in the domain of artificial cognitive systems and distributed human–machine systems (DHMS) is limited only by the creative imagination. Accordingly, with sustained scientific leadership, there is reason for optimism about the continued development of (1) specialized artificial cognitive systems that emulate specific aspects of human performance and (2) DHMS, whether through approaches that are faithful to cognitive neurophysiology, or through some mix of engineering and studies of human intelligence, or by combining the respective strengths of humans and automation working in concert. Researchers are addressing the limitations that made earlier systems brittle by exploring ways to combine human and machine capabilities to solve problems and by modeling coordination and teamwork as an essential aspect of system design.

Framework Relevance: Clearly the emphasis on a systems engineering approach is a key feature of the Framework. The following committee statement also supports the Framework's emphasis on risk-informed surety engineering practices based on maturity identification and quantifiable evidence that would improve the measurement (performance and progress) indicators and consideration of ELS issues such as unintended use.

> *"Research in artificial cognitive systems and distributed human–machine systems has been hampered by unrealistic programs driven by specific, short-term DOD and intelligence objectives. A second problem is the inadequacy of current approaches to metrics. Resolving this problem would enable meaningful progress. Finally, the study of ethical issues related to the design and deployment of distributed human–machine systems is virtually in its infancy. This is deplorable given the great potential of such systems for doing good or harm."*

Report Cultural and Ethical Implications: The report provided significant emphasis on cultural and ethical implications of cognitive system research and applications. Statements include:

> *"Research is enhancing understanding of how culture affects human cognition, including brain functioning, and is even suggesting a link between culture and brain development. The U.S. military is placing greater emphasis on cultural awareness training and education as a critical element in its strategy for engaging in current and future conflicts. Military conflicts will increasingly involve prolonged interaction with civilian populations in which cultural awareness will be a matter of life and death and a major factor in outcomes."*

> *"The brain is viewed as the organ most associated with personal identity. There is sure to be enormous societal interest in any prospective manipulation of neural processes."*

A normative reference for ethical treatment of human participants in biomedical research is the World Medical Association's Declaration of Helsinki (DOH).

[43] T. Trucano et al. "R&D for Computational Cognitive and Social Models: Foundations for Model Evaluation through Verification and Validation," SAND2008-6453, September 2008.

> *“Although the international community largely accepts and respects the DoH, data on compliance by individual states are not available.”*

Other such guidelines are identified in the report followed by the statement:

> *The various guidelines show consensus on some main beliefs including that the research must be reviewed from an ethics standpoint before it is conducted; that the research must be justifiable and contribute to the well-being of society in general; that the risk-benefit ratios must be reasonable; that informed consent or voluntariness is needed; that there is a right to privacy; that accurate reporting of data is obligatory; and that inappropriate behaviors must be reported.*

Framework Relevance: The cultural and ethical implications addressed in the report indicate the correctness of the Framework’s emphasis on the engineering integration of ELS issues within the cognitive system research, development, and product implementation.

APPENDIX D - COGNITIVE SYSTEMS AND ELS RISK RESEARCH

D.1 Cognitive Neuroscience Inspired Models

D.1.1 Specification Model Concepts

The Specification Model for cognitive systems can be considered at four conceptual levels of analysis (Sun 2008[44]): 1) inter-agent, involving environmental elements of the Specification Model, 2) agent, involving environmental and psychological elements, 3) intra-agent, involving psychological and physiological elements, and 4) substrates, involving physiological elements of the Specification model. Here, an agent is the generic reference to the entity type being modeled (i.e. human). Implicit in all of these levels of analysis is that the models are embodied, that is, the models are considered to be embedded in a body that is located in and interacts with an environment. This environment can be normal, abnormal or hostile. The cognitive model, bodies, and environments may be simulated or implemented physically. When dealing with software implementations of these models, standard software engineering surety concerns apply.

The first level of analysis (inter-agent) is associated with "social and cultural" processes, involving groups of agents, collective behavior, interactions between individuals and groups of agents, and interactions with their environment. The second level (agent) is associated with the "psychology" of the agent, involving individual behaviors, its knowledge, beliefs, perceptions and actions, and learning. The third level (intra-agent) is associated with the "components" of the cognitive model, including cognitive architectures, modular neural networks, function and structure analysis, symbolic computation, computational languages, hierarchical analysis, and again learning. Finally, the fourth level (substrates) of analysis is associated with the "physiology" or implementation of the components, including neurobiological mechanisms, neuroscience, neural processes, gene expression, ion channel mechanisms, synaptic structures, simulation languages, and again learning.

The following describes an approach to cognitive modeling focused at the intra-agent level of analysis, dealing with the Specification Model primarily at the components level involving psychological and physiological elements. It introduces an appropriate mathematics and connects it to the underlying theoretical foundations of cognitive systems[45]. Other normative references for this approach are[46,47,48].

D.1.2 Design Model Concepts

Design models of episodic memory in humans are usually based on a recurrent layered neural network structures isomorphic to the understood functional anatomy of the human brain (K. Norman et al 2008[49]). The input layer to this class of model is frequently taken to be the "top end" of Medial Temporal cortex, with multiple layers for the entorhinal cortex, the dentate gyrus and tightly coupled layers representing he CA1 and CA3 sub-regions of the hippocampus. The

[44] R. Sun (Editor), The Cambridge Handbook of Computational Psychology, Cambridge Press, 2008.

[45] M. Healy, T. Caudell, Ontologies and Worlds in Category Theory: Implications for Neural Systems, Axiomathes, 2006.

[46] E. Kandel, J. Schwartz, T. Jessell Principles of Neural Science, McGraw-Hill, 2009.

[47] P. Churchland, Brain-wise, MIT Press, 2002.

[48] L. Barsalou, Perceptual Symbol Systems, Behavioral & Brain Sciences, 1999.

[49] "Computational Models of Episodic Memory", K. Norman, G. Detre, and S. M. Polyn, Ch.7 in The Handbook of Computational Psychology, R. Sun Editor, Cambridge Press (2008).

models have wide spread use of recurrent or feedback connections to support the possibility of neurodynamical processes like “memory attractors”. These attractors provide a way that the system can lock on to a particular episode for short periods of time and to perform pattern completion. Hebbian learning algorithms are used to capture and reinforce associations within episodes. The attractors formed within the layer structure will function to perform pattern recognition and to detect novel, non-matching patterns. This also gives the model the capacity to recall the patterns and to use them to make conceptual predictions of what is to come. This class of model is frequently used as a part of a larger more functional cognitive model, and may share properties with other memory systems in the brain.

D.1.3 Mathematical Modeling Concepts

Category theory is the branch of mathematics concerned with pure structure. In recent work, it has been applied to the formalization of the underlying structure of knowledge organized into ontologies for computational systems, including the exploration of ontologies for comprehensive, unambiguous, system-language-neutral knowledge representation. This section introduces our application of this mathematics: The incremental acquisition and representation of ontologies by adaptive neural networks. Category theory can be applied to ontologies for understanding the many possible worlds that are internalized implicitly in the computations of adaptive neural networks. This theory constitutes a scientific theory of brain structure/function, providing a fundamental mathematical description of how these components function and interact.

From the earliest writings on the subject, investigators in logic and the semantics of computation have sought an accurate understanding of the implicit meaning of neural computations. Presumably, when a well-designed network adapts its connection weights, it is effectively modifying its responses to input stimuli to improve its future response according to some criterion. Information gained from the input data becomes internalized in the form of connection weight modifications, which affect the future response of the network to its inputs. What is this information; is it possible to understand it as knowledge expressible in a human-understandable form? This question is often addressed by attempting to decode the adapted connection-weight values as logical rules that the network is supposed to have learned from its input stream. Formal-logic-like languages are sometimes used for this, to allow declarative statements such as IF-THEN rules to be expressed without ambiguity and sometimes with full mathematical rigor. Intuitively, the ability of any computational system to manipulate data in a systematic way is a manifestation of the knowledge represented in the system’s design and in its store of already-processed data. The use of a mathematical language to understand the knowledge content of a computational system is called mathematical semantics.

Cognitive neuroscientists seek to understand the relationship between structure and function in the brain – the semantics of neuron/synapse organization. One of their significant findings is that neurons and their synaptic connections are organized on a larger scale into a system of interconnected functional modules. Each module is associated with one or more sensory modalities, motor control, planning, and the control of working memory, and a module implicated in self-referential processing has been tentatively identified in humans.

One way to regard this is to assign to each module a system of knowledge that appears to describe the functionalities with which it has been associated. For example, the recent ‘‘what/where’’ model of the visual pathway from the retina to other areas of the brain asserts that the pathway bifurcates. The spatial layout of objects in the visual field appears to be extracted in successive processing stages along a pathway from the occipital to the posterior parietal lobe in the cerebral cortex (the dorsal path). In parallel with this, a pathway from the occipital lobe through the temporal lobe (the ventral path) appears to form semantic object representations in successive stages. The representations begin as simple sensory features and eventually reach the complexity of scenes and events near the juncture of parietal, temporal and occipital lobes, where multi-sensor representations of scenes and events appear to be formed. Connection pathways between regions help organize the more complex object, scene, and event representations among the modalities. Apparently, they do this by re-uniting the spatial and semantic information, now broken down into iconic representations manipulable in a flexible working memory system. The working memory system is a set of processes involving functional regions and their interconnections that organize, store and recall information from synaptic memory that is associated with current experience.

This description of the findings of cognitive neuroscience suggests a view of the brain as a knowledge-manipulation system that acquires information and forms separate representations of it, beginning with sensor-related knowledge such as visual form, auditory form, and spatial location. The storage process is more than a simple filing-away of data; the flexible use of data, involving creativity in many organisms, suggests that what is stored is a ‘‘internal model’’ of the world. This model is many-faceted, capable of representing a wide variety of situations that can be associated with simultaneous inputs from several sensors.

How does this multi-faceted view of knowledge representation in the brain relate to the expression through the organism’s behavior of a single, unified system of knowledge? After all, an organism does not jump between visual, auditory and other knowledge representations, applying them one at a time in a disconnected fashion, for such incoherence would lead to disaster. The categorical mathematical semantic model explains the interactions of the modular knowledge representations in a multi-module network through interconnection pathways as a unifying of representations. In this unification, the whole network acts as if there were a single knowledge system guiding its behavior, a key property that is called “knowledge coherence”.

In summary, an ontological knowledge structure and its incremental representation through adaptation in neural network architectures can be formalized in an appropriate and fundamental mathematics: category theory. The categorical constructs used to model the representation of concepts and their inter-relationships suggest cognitive model architectural structures and their properties. The result is a set of principles for specification of a cognitive model design that applies to learning and to the combination of information from multiple sensors in multi-module architectures. These principles are being applied to neural-based cognitive model design and analysis for evaluation. By providing a mathematical vehicle for associating a hierarchy of concepts with a multi-modular neural architecture and explicating the incremental learning of both more abstract and more specific concepts with re-use of existing conceptual knowledge, the theory has a natural role as a fundamental theory for exploring knowledge representation in neural networks, assisting in the design of more realistic cognitive models.

D.2 Ethics, Legal, Societal Issues and Potential Risks

The discipline of surety engineering offers a rigorous and systematic approach to the identification and analysis of the spectrum of risks potentially triggered by the development and dissemination of cognitive systems and neuro-technologies. A substantial number of these risks are likely to be in the spheres of law and ethics, but it can be difficult to identify those risks with specificity, particularly in the very early stages of the development of a novel technology. The surety methodology outlined in this report provides a framework in which technology developers can be prompted to anticipate legal and ethical concerns associated with their work and to do so beginning with the basic research stage of a project and continuing on with increasingly detailed analysis as a product or process is offered to commercial and government customers.

The potential reach of cognitive systems technologies is very broad, and the ethical and legal implications of computational and/or biologically-based models that aspire to replicate or simulate human thought processes are substantial. As exemplified in Descartes' famous dictum "I think therefore I am," the Western tradition places the thinking brain at the core of human identity. The use of technology to penetrate the inner workings of the mind – by reading the mind, copying it, enhancing it, or degrading it – implicates deeply held convictions about individual control over this most personal domain.

Important legal and ethical questions cut across at least six general areas:

(1) Responsible science: the responsibility of cognitive technology developers to assure the safety and reliability of cognitive systems, to disclose all known limitations, and to avoid exaggerated claims of the capabilities of new products;

(2) Privacy: the individual's right to control access to his or her specific thoughts and to any measures or representations (whether biological or computational) of his or her cognitive capabilities;

(3) Informed consent and control: the individual's right to complete disclosure of the risks and benefits of any technology that purports to measure, record, analyze, model or intervene in cognitive activity, coupled with the right to decide whether to accept the use of the technology;

(4) Public dialogue: the shared responsibility of technology developers and relevant communities ("the public"), including government representatives, to exchange information about the broader impacts and risks of cognitive systems and to collaborate in developing appropriate strategies for governance of these potentially transformative technologies;

(5) Human enhancement: the controversial prospect of an expanding set of options for using technology to improve mental performance and to extend human cognitive capabilities, viewed by some as impermissible interference with "nature" and by others as consistent with a moral imperative to improve the human race; and

(6) Security: the recognition of the potential for intentional misuse of cognitive technology (such as unauthorized use of private information regarding individual cognitive data or intervention in an individual's cognitive activity) and implementation of appropriate protections and "countermeasures."

In addition, special considerations may apply to certain cognitive systems, known as “cognitive models,” which are derived from an individual’s personal cognitive data, but designed to be able to operate in contexts that are independent of the prototype individual. The potential to create independently operating representations of human individuals raises questions of ownership and personal identity – are the data and capabilities of a cognitive model the property of the prototype person, the developer of the cognitive model, the employer who supported the creation of the model? Is there an obligation to acknowledge the contribution of the prototype person or to ensure that the model is an accurate representation of a person who is identified as the basis for the model? Given that a cognitive model may also be designed to evolve over time in response to various environmental inputs, how do these property and identity questions change as the connection between the model and the prototype person arguably becomes more attenuated? Ultimately, might a model acquire moral standing of its own?

The prospect of a cognitive model operating and evolving independently raises unique ethical and legal questions about the responsibility for decisions made and actions taken by the model. Imagine a cognitive model that simulates the mental processes of an expert in a particular domain – training fighter pilots, for instance. If a pilot receives training from the model and accidentally shoots down another aircraft in an incident of “friendly fire,” how should we assess the relative fault of the expert trainer who provided inputs for the model, the designer of the model, and the pilot? Is there any sense in which the model itself can be held accountable? Would it be plausible to use a surety approach to build basic ethical precepts into such a cognitive model? Where a cognitive model is designed to evolve in response to multiple human and environmental inputs, can a “black box” be incorporated into the model and used to help establish a trail of accountability?

The development of cognitive systems technologies, including cognitive models, is currently at a very early stage, but many of these legal and ethical issues are beginning to be identified, with the caveat that actual technologies may not deliver on all the promises suggested by preliminary research. The remainder of this section elaborates on the six cross-cutting legal/ethical issues areas identified above, with examples and discussion of concerns that can be analyzed and addressed using surety methods.

D.2.1 Responsible Science

One of the significant challenges in developing new technologies is the assessment of safety and reliability of products with unprecedented capabilities. The nature of a new technology is such that the associated risks – to research subjects, potential consumers, humans, the environment, and society at large – are often difficult to foresee. At the same time, the legal system, through rules of tort liability, places principal responsibility for anticipating and managing such risks at the feet of those who develop and market these technologies. The rationale for this approach is that those who design and sell a product are in the best position to know how to create a safe product and to spread the risk (generally through liability insurance) of malfunctions and unanticipated events. For some types of products, in the pharmaceutical industry for example, government review and approval of safety and efficacy (using evidence from large scale clinical trials) are also required in advance of marketing a new drug or device to the general public. Other consumer protection and product liability laws, which vary across jurisdictions, govern a manufacturer’s obligations to disclose known risks, to avoid negligent or fraudulent misrepresentations concerning a product, and to compensate individuals who can show that they have suffered harm caused by a manufacturer's failure to take reasonable preventive steps.[50]

The technology developer's professional ethics likewise emphasize the responsibility to safeguard the public welfare. For instance, the Institute for Electrical and Electronics Engineers (IEEE) contains the following commitment in its Code of Ethics:

> We, the members of the IEEE, in recognition of the importance of our technologies in affecting the quality of life throughout the world, and in accepting a personal obligation to our profession, its members and the communities we serve, do hereby commit ourselves to the highest ethical and professional conduct and agree: …To accept responsibility in making decisions consistent with the safety, health and welfare of the public, and to disclose promptly factors that might endanger the public or the environment; (emphasis supplied).[51]

In addition, when the research and development of a new technology involves testing on human subjects, ethical research practices require careful assessment of the scientific merit of the research involved, as part of the analysis of the relative risks and benefits performed in connection with review by an institutional review board.[52]

D.2.2 Privacy

Privacy is a complex and multi-faceted concept, incorporating a broad range of subjective individual and cultural values. Protection of privacy generally involves several different areas of concern, including privacy of personal information and privacy of personal physical space, as well as privacy of decision-making without interference from government or third parties. Proprietary rights of ownership and control over one's own unique personality can also come under the umbrella of the concept of privacy.[53]

The U.S. legal system recognizes rights of privacy in various contexts – through privacy protections in the Constitution, as well as state and federal statutes, which mandate special treatment of certain personal information and personal spaces. These constitutional privacy protections generally restrict intrusions into private matters by government actors and enshrine what has long been called "the right to be let alone."[54] The First Amendment of the Constitution protects the rights of individuals to express themselves, through freedom of speech and through association with groups of their choosing. The little-noted Third Amendment prohibits the government from requiring citizens to quarter soldiers in their homes, a recognition of private space in the home that dates back to revolutionary war days. Similarly, the Fourth Amendment protects the "right of the people to be secure in their persons, houses, papers and effects against

[50] *See* "Products Liability Law: An Overview," Legal Information Institute, Cornell University Law School, available at http://topics.law.cornell.edu/wex/products_liability

[51] IEEE Code of Ethics (2006), available at http://ethics.iit.edu/codes/coe/inst.electrical.electronics.engineers.2006.html

[52] *See* Freedman, Benjamin. "Scientific Value and Validity as Ethical Requirements for Research: A Proposed Explication," *IRB: Ethics and Human Research*, Vol. 9, No. 6 (Nov. – Dec. 1987), pp.7-10

[53] A. Allen, "Genetic Privacy: Emerging Concepts and Values," p. 33, in Genetic Secrets, ed. by Mark Rothstein (Yale University Press, 1997).

[54] *See* "Privacy Law: An Overview," Legal Information Institute, Cornell University Law School, available at http://topics.law.cornell.edu/wex/privacy. The constitutional "right to be left alone" was enunciated by Supreme Court Justice Louis Brandeis in a well-known dissenting opinion in the case of Olmstead v. United States, 277 U.S. 438 (1928), available at http://www.law.cornell.edu/supct/html/historics/USSC_CR_0277_0438_ZD.html

unreasonable searches and seizures" by government authorities. This protection limits government intrusions that interfere with a person's "reasonable expectation" of privacy. The Fifth Amendment provides for a privilege against self-incrimination, again protecting a zone of personal privacy, although this zone is a narrow one, protecting only "testimonial" evidence while allowing the government to compel a person to provide physical evidence such as a blood sample. Some neuro-ethics scholars are beginning to ask whether neuro-technology will be used to "search and seize" the contents of a suspect's brain or whether witnesses might be compelled to undergo brain scans in spite of the privilege against self-incrimination.[55]

Concerns about privacy extend beyond the prospect of intrusions by government entities; many current privacy concerns arise from the power of corporations to gather, use and sell private information in contexts ranging from insurance to employment to targeted marketing. The United States has no general statutory law governing privacy; however Congress has enacted laws governing privacy in specific domains – for example, health care records (Health Insurance Portability and Accountability Act), educational records (Family Educational Rights and Privacy Act), and financial information (Fair Credit Reporting Act). The new Genetic Information Non-Discrimination Act is scheduled to take effect in November 2009 and expressly prohibits insurers or employers from making improper use of genetic information to deny insurance or employment to individuals. This past July, the Senate Commerce Committee held a hearing to assess the need for comprehensive privacy legislation to set standards for the collection of personal data on-line. The purpose of the hearing was to address concerns that "tracking individuals' Internet activity and gathering information from online users violates their expectations of privacy. Individuals often are unaware what information is being collected about them, how it is being used and to whom it is disseminated."[56] Although no legislation was proposed, several large internet companies were supportive of greater regulation, while the Federal Trade Commission defended the current approach of self-regulation, under which companies are encourage to disclose to users that their data is being collected and to offer users the chance to opt out of the data collection.[57]

Many states have statutes and provisions in their state constitutions that address a right to privacy in general, as well as specific kinds of privacy.[58] So-called "privacy torts" also impose civil liability for acts that can be characterized as privacy violations: intrusion into a person's private affairs, public disclosure of embarrassing facts, placing a person in a false light before the public, or appropriation of a person's name or likeness. Use of a cognitive model of an individual to represent that person in ways he or she finds objectionable – or to profit commercially by selling the model – could give rise to liability, particularly if the model is developed or used without the input and agreement of the individual. In California, for example, a state statute expressly

[55] S. Tovino, "Functional Neuroimaging and the Law: Trends and Directions for Future Scholarship," *AJOB Neuroscience* Vol. 7, No. 9, pp. 44-56. (Sept. 2007)

[56] Privacy Implications of Online Advertising, Hearing of the Senate Commerce Committee, July 9, 2008, http://commerce.senate.gov/public/index.cfm?FuseAction=Hearings.Hearing&Hearing_ID=e46b0d9f-562e-41a6-b460-a714bf370171

[57] S. Hansell, "Senators Weigh Possible Rules for Advertising and Online Privacy." *New York Times* (July 9, 2008).

[58] The state of California, for example, provides for each citizen's "inalienable right" to pursue and obtain "privacy" in its state constitution; specific state laws govern privacy in various arenas from electronic surveillance to health records to identity theft. *See* California Office of Information and Privacy Protection website at http://www.oispp.ca.gov/consumer_privacy/laws/

prohibits the unauthorized commercial use of a person's name, voice, signature, photograph or likeness, a right which continues for 70 years after a person's death.[59]

D.2.3 Informed Consent and Control

An important dimension of the legal and ethical concerns in the area of brain-based technology is informed consent. The concept of informed consent became prominent following the post World War II Nuremberg trials of Nazi doctors who had conducted experiments on concentration camp victims who were powerless to object. Following the trial, the panel of American judges issued what is know as the Nuremberg Code, a seminal document in the field of human subject research ethics, which begins with this principle: "The voluntary consent of the human subject is absolutely essential."[60] Subsequent articulations of research ethics principles likewise emphasize the need for informed consent,[61] and federal law governing research in federally funded institutions also makes this requirement a cornerstone of the regulation of human subject research.[62]

The required elements of informed consent to a medical intervention or to participation in research are:

1. Competence;
2. Disclosure;
3. Understanding;
4. Voluntariness; and
5. Consent.[63]

If these requirements are not met – for example, if a patient does not receive full disclosure of relevant risks and benefits of a proposed procedure – the physician may be subject to a legal claim of battery. If these requirements are not met by a researcher using human subjects, federal funding of research at his or her institution may be jeopardized. To the extent that cognitive systems and neuro-tech research involves the use of human subjects, these informed consent requirements specifically apply, along with the federal regulations mandating review and approval of the research by an institutional review board.

D.2.4 Public Dialogue

The concept of informed consent can also be used to characterize a technology developer's ethical obligations to prospective customers and, especially where the new product has broad ramifications, to the public at large. It can be argued that these customers and the wider society

[59] State Right to Publicity Laws, National Conference of State Legislatures, available at http://www.ncsl.org/programs/lis/privacy/publicity04.htm; California Civil Code Section 3344, 3344.1, available at http://law.onecle.com/california/civil/3344.html

[60] Nuremberg Code, *Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law No. 10, Vol. 2, pp. 181-182.*. Washington, D.C.: U.S. Government Printing Office, 1949, available at http://ohsr.od.nih.gov/guidelines/nuremberg.html

[61] See The Belmont Report, National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (April 18, 1979), available at http://ohsr.od.nih.gov/guidelines/belmont.html and WORLD MEDICAL ASSOCIATION DECLARATION OF HELSINKI: Ethical Principles for Medical Research Involving Human Subjects (last revised 9/10/2004), available at http://ohsr.od.nih.gov/guidelines/helsinki.html

[62] Federal Policy for the Protection of Human Subjects, 45 C.F.R. 46, available at http://ohsr.od.nih.gov/guidelines/45cfr46.html#top

[63] Beauchamp & Childress, Principles of Biomedical Ethics (5th ed. 2001), p. 79.

are in some sense the “human subjects” in an ongoing collective experiment with the potential to change society – and change conceptions of human nature – in fundamental ways. To facilitate informed decision-making about the risks and benefits of transformative technologies, technology developers can use their expertise to educate and inform others about the issues they perceive. By improving public understanding and promoting full disclosure of what is known about a new technology, technology proponents can provide a meaningful opportunity for individuals and policymakers to base decisions to accept or reject unprecedented changes on an accurate understanding of the underlying science.

The Universal Declaration on Bioethics and Human Rights, adopted by UNESCO (United Nations Educational, Scientific and Cultural Organization) in 2005, refers specifically to the need for broad-based dialogue and decision-making about bioethical issues:

> Persons and professionals concerned and society as a whole should be engaged in dialogue on a regular basis.
>
> Opportunities for informed pluralistic public debate, seeking the expression of all relevant opinions, should be promoted.[64]

In the United States, one prominent forum for such public debate is the President’s Commission on Bioethics, which has conducted several meetings on topics in neuroethics[65] and has issued a report documenting concerns about human enhancement, including cognitive enhancement, entitled Beyond Therapy.[66]

In addition, a variety of professional organizations have engaged in scholarly discussions and debates about the emerging discipline of neuroethics. The *American Journal of Bioethics* has recently expanded to include regularly published issues on ethics and neuroscience, and another new journal, entitled *Neuroethics,* was introduced by Springer Press in 2008.[67] Scholarly work has also focused on legal issues, particularly those triggered by brain-scanning technologies.[68] In 2007, the MacArthur Foundation made a $10 million grant to establish the Law and Neuroscience Project, an interdisciplinary effort to examine the intersection between neuroscience and the courtroom and to provide education and outreach to judges and others who influence the evolution of law and policy relating to neuroscience.[69] Mainstream media has also begun to publish articles that examine a wide range of topics linking neuroscience findings to social, legal and ethical concerns in articles with titles like “The Brain on the Stand,” describing concerns that “the use of brain-scanning technology as a kind of super mind-reading device will threaten our privacy and mental freedom, leading some to call for the legal system to respond with a new concept of ‘cognitive liberty.’” [70]

[64] Universal Declaration on Bioethics and Human Rights, United Nations Educational, Scientific and Cultural Organization (2005) available at http://portal.unesco.org/en/ev.php-URL_ID=31058&URL_DO=DO_TOPIC&URL_SECTION=201.html

[65] Transcripts available at http://www.bioethics.gov/topics/neuro_index.html

[66] On-line copy available at http://www.bioethics.gov/reports/beyondtherapy/index.html

[67] Several anthologies of scholarly essays have also been published, including Neuroethics: Defining the Issues in Theory, Practice and Policy, ed. by J. Illes (Oxford University Press 2006) and Defining Right and Wrong in Brain Science, ed. by Walter Glannon (Dana Press 2007).

[68] *See, e.g.*, Neuroscience and the Law: Brain, Mind and the Scales of Justice, ed. by B. Garland (Dana Press 2004).

[69] The Law and Neuroscience Project website can be found at http://www.lawandneuroscienceproject.org/

[70] J. Rosen, “The Brain on the Stand,” *New York Times* (March 11, 2007).

Most of these public discussions concern the uses of neuro-imaging devices; few have addressed the possibilities of other brain-based technologies that go beyond so-called "mind-reading" to the kind of "mind-mimicking" entailed in creation of cognitive models or the kind of "mind-altering" that may be associated with brain implants or other developing technologies. Although we have had extensive public and policy debate on the risks and benefits of "mind-altering" psycho-pharmaceuticals,[71] many questions remain about these and other possibilities for mind control.

From a surety perspective, good faith efforts to foster the expansion of the public dialogue to include cognitive systems are important risk mitigation measures. Examples abound of novel technologies that caused public controversy, often to the surprise of the scientific community; stem cell research and GMO foods are examples that come readily to mind. To reduce such public relations risks, careful consideration should be given to mechanisms for stakeholder dialogue, with broad and diverse participation from affected communities. Examples of such structured conversations are being studied in connection with the introduction of nanotechnology in the United Kingdom, where citizen groups have taken on the role of "nano-juries" to review ethical issues.[72] In the United States, the NanoFutures Project at the Center for Nanotechnology in Society is using a web-based conversation (with participants invited to post comments and revisions to futuristic scenarios) to conduct "an experiment in creating social engagement around anticipatory governance of nanotechnology."[73] Although this project is focused on nanotechnology, several of the scenarios under consideration involve cognitive technologies – a brain chip featuring a data feed that puts information in the brain while the user is resting and an optical implant enabling magnification, infra-red visualization and night vision.

D.2.5 Human Enhancement

The development and conceptualization of technologies with the potential to significantly extend and enhance the capabilities of the human brain has spurred wide-ranging debates about the ethics of cognitive enhancement.[74] Advocates for what is known as "transhumanism" embrace a vision for the future in which technology is used to improve upon and even "transcend" human nature – by offering increased longevity, expanded physical and intellectual capabilities, and more control over our moods and mental states.[75] Proponents of a "bio-conservative" point of view, on the other hand, argue that emerging biotechnologies may alter human nature – the "stable human essence" that individuals have in common – and undermine fundamental human dignity.[76]

Although many of the futuristic scenarios imagined in this debate are far from being realized, some kinds of cognitive enhancements are already available and in use. Drugs to limit mental

[71] See, e.g., Beyond Therapy: Biotechnology and the Pursuit of Happiness, A Report by the President's Council on Bioethics (2003), available at http://www.bioethics.gov/reports/beyondtherapy/index.html

[72] NanoJury UK, at http://www.bbsrc.ac.uk/society/dialogue/activities/nanojury.html

[73] http://www.brainery.net/nanofutures/index.html

[74] M. Farah, et al., "Neurocognitive Enhancement: What Can We Do and What Should We Do?" *Nature Reviews Neuroscience*, Vol. 5, Issue 5 (May 2004) pp. 421-425.

[75] N. Bostrom,"In Defense of Posthuman Dignity," *Bioethics*, Vol. 19, No. 3 (2005) pp. 202-214; available at http://www.nickbostrom.com/ethics/dignity.html

[76] F. Fukuyama, Our Posthuman Future: Consequences of the Biotechnology Revolution (Picador, 2002), pp. 217-218.

fatigue and reduce human error are being studied in military settings.[77] A recent on-line poll conducted by the journal *Nature* found that one in five of its readers who responded had used drugs – methylphenidate (Ritalin), modafinil (Provigil), or beta blockers – for non-medical, cognition-enhancing purposes.[78] In addition, pharmaceuticals for improving memory, as well as dampening memory (of traumatic experiences, for example), are under development. Also on the horizon are non-pharmaceutical enhancement technologies – such as brain implants (already used to alleviate symptoms of Parkinson's disease) and non-invasive brain stimulation using transcranial magnetic stimulation (TMS).

Still other methods for extending human cognitive capabilities are being developed through the use of information technology and the growing research in cognitive systems. In one widely cited report generated by a conference sponsored by the National Science Foundation, participants articulated a vision for research combining cognitive science with nanotechnology, biotechnology and information technology to achieve such far-reaching goals as:

- Fast, broadband interfaces directly between the human brain and machines will transform work in factories, control automobiles, ensure military superiority, and enable new sports, art forms and modes of interaction between people;
- Comfortable, wearable sensors and computers will enhance every person's awareness of his or her health condition, environment, chemical pollutants, potential hazards, and information about local businesses and the like;
- The human body will be more durable, healthier, more energetic, easier to repair, and more resistant to many kinds of stress, biological threats, and aging processes.[79]

Additional risks associated with these emerging cognitive technologies include the likelihood that access to enhancements will be unequal, exacerbating existing socioeconomic divisions, and the possibility that the availability of enhancements will increase pressure to do whatever it takes to compete in cognitive endeavors.

D.2.6 Security

Many of the legal and ethical issues associated with cognitive systems and neuro-technologies concern the responsibility to preserve the security of data reflecting deeply personal information about the unique workings of a person's brain. As discussed above, the collection of this cognitive information, whether through imaging of neurons or analysis of behavioral manifestations of cognition, can lead to a violation of an individual's privacy rights, especially if it is done without the knowledge of that individual. Once the cognitive data is collected, the ongoing use of that data is likely to pose additional privacy and security risks. These risks arise from the processes and personnel used in recording, storing, analyzing and systematizing the information in various ways, including the possible creation of individualized cognitive models, linked to specific identifiable persons.

[77] J. Moreno, "Juicing the Brain: Research to limit mental fatigues among soldiers may foster controversial ways to enhance any person's brain," *Scientific American Mind*, Vol. 17, Issue 6 (Dec. 2006), pp. 66-73.
[78] B. Maher, "Poll Results: Look Who's Doping," *Nature*, Vol. 452 (April 10, 2008)
[79] M. Roco and W. Bainbridge, eds. Converging Technologies for Improving Human Performance: Nanotechnology, Biotechnology, Information Technology and Cognitive Science (NSF/DOC-sponsored report, 2002), p. 5.

At every step of this process, there is a risk that cognitive data, which may be highly personal (reflecting an individual's emotional vulnerabilities, for example), could be disclosed or misused to the individual's detriment. Imagine, for instance, the prospect of identity theft that goes beyond theft of personal data to theft of a digital representation of someone's unique personality or skills. These security risks are heightened by the prospect of technologies that will not only "read" minds (and provide the data to governments, employers, private detectives, to name a few possibilities) but may also intervene in and alter a person's cognitive processes.

Under current laws enacted in response to increased prevalence of identity theft, entities collecting and holding personal financial information are responsible for disclosing data security breaches to affected consumers; at least 44 states have enacted legislation requiring notification of security breaches involving identifiers such as Social Security numbers and driver's license numbers.[80] The purpose of these disclosure laws is to enable consumers to take steps to prevent identity theft. Similar disclosure obligations would likely apply to security breaches of cognitive data.

To prevent or minimize the possibility of such security breaches, developers of cognitive technologies should take seriously the responsibility of including safeguards that protect the cognitive privacy through techniques such as encryption of personal data. Additionally, attention needs to be given to the policies and procedures that govern the secure storage of cognitive information collected from individuals, as well as the development of consistent policies governing the retention, destruction and permissible future uses of such data. Recent debates about genetic privacy in connection with the rapid expansion of DNA testing and databases – for purposes ranging from medical research to criminal law – have highlighted public concerns about the existence of repositories of sensitive personal data that may be used in ways that were not intended or disclosed at the time of data collection.[81] Clear rules, specified in advance and agreed to by individuals providing access to their cognitive data, will mitigate some of the risks associated with these informational privacy concerns.

One example of such prospective disclosure can be seen in the context of another technology with the potential to "track" individual behavior – event data recorders, or "black boxes," in cars.[82] These data recorders, found in millions of newer vehicles, can capture and preserve information about driver speed, brake usage, airbag release, seat belt usage, and many other computerized vehicle functions. This information can then be used to analyze accidents, determine fault, and help adjudicate product liability claims. Although car companies and federal regulators view this information as beneficial for investigation of accidents and development of safety improvements, some privacy advocates have objected to this technology. Since 2004, 12 states have enacted laws relating to event data recorders, including laws requiring that the presence of the device be disclosed to a vehicle's purchaser and limiting the permissible uses of the information collected. The federal National Highway Traffic Safety Administration has also issued a rule that will require automakers (beginning with model year 2011) to disclose to new

[80] State Security Breach Notification Laws, National Conference of State Legislatures, available at http://www.ncsl.org/programs/lis/cip/priv/breachlaws.htm

[81] Despite these concerns about ownership and control of a person's unique genetic material, three different courts have found that individuals do not retain property rights in their genes. For discussion of these cases and related issues, see Rao, Radhika, "Genes and Spleens: Property, Contract, or Privacy Rights in the Human Body?" *Journal of Law, Medicine and Ethics* (Fall 2007).

[82] M. Wald, "Does Your Car Have a Spy in the Engine?" *New York Times* (Oct. 27, 2004).

car buyers whether an event data recorder has been installed in a vehicle.[83] These policies limit the use of black box "tracking" by requiring disclosure to the consumer but not prohibiting the technology altogether. In some cases, specific uses of similar technologies may be further restricted. For example, New York and California have adopted laws prohibiting rental car companies from using electronic surveillance to impose fees or charges on the renter of a vehicle.

Finally, an additional dimension of security that should be addressed is national security. Some of the technologies that are under development in this "neuro-tech" arena involve military applications with implications for national security. Indeed, one of key supporters of research in this area is the Defense Advanced Research Projects Agency (DARPA), which is actively pursuing projects to augment human cognition for use in military settings.[84] The potential military advantage associated with such innovations might well be compromised by failure to appropriately protect not just individual data but also project data.

In another example of military use of neuro-technology, Army neuroscientists are developing "thought helmets" that will capture a soldier's brainwaves and relay them from one soldier to another to create an avenue for silent and secure communication among troops.[85] If these "thought helmets" are to serve their intended purpose, security features will be necessary to prevent eavesdropping on these cognitive communications, and it may be that development of security measures for such projects will provide needed technology for privacy protection in the civilian context as well.

Additional national security issues arise from the need to understand the technological capabilities of other state and non-state actors whose interests may not be congruent with those of the United States. A recent report by the National Academy of Sciences, entitled "Emerging Cognitive Neuroscience and Related Technologies," recommends that U.S. intelligence capabilities in this area be expanded, in recognition of the fast pace of international research in this area and of the vast potential for cognitive systems and related technologies to become powerful military tools requiring development of countermeasures.[86]

The ethical issues that arise in connection with national security and military necessity are myriad – civilian casualties, escalating arms races, military ethics and so on. These issues are beyond the scope of this document but will likely need to be addressed by those who work in this arena.

D.2.7 Example of Cognitive System ELS and Technical Risk Perceptions: fMRI

Example Study: Legal, Ethical and Policy Issues in Response to the Current and Projected Uses of Brain Scanning Technology for Deception Detection and Other Possible Forms of "Mind-Reading"

Because many cognitive systems technologies are currently in the research and development phase, predictions about their legal and ethical implications necessarily contain an element of

[83] See 2007 Privacy Legislation Related to Event Data Recorders ("Black Boxes") in Vehicles, National Conference of State Legislatures, available at http://www.ncsl.org/programs/lis/privacy/blackbox07.htm

[84] See J. Moreno, Mind Wars: Brain Research and National Defense (Dana Press 2006).

[85] M. Thompson, "The Army's Totally Serious Mind-Control Project," TIME (Sept. 14, 2008).

[86] Emerging Cognitive Neuroscience and Related Technologies, National Academies Press (2008) http://www.nap.edu/catalog/12177.html

speculation. To shed light on possible societal responses to a variety of potentially new kinds of brain-based systems, this case study turns to the existing neuro-technology of brain imaging (fMRI or functional magnetic resonance imaging). Developed as a medical diagnostic tool, neuro-imaging is now being used in other contexts, from neuroscience research to lie detection to assessments of criminal responsibility. These new applications – actual and projected – have triggered a broad spectrum of reactions, especially in connection with the potential for using neuro-imaging to "read" minds. In one recent case, a brain scan was recently accepted by a court in India as proof that the accused had "experiential knowledge" that could only have come from committing the murder in question and that was therefore the basis for conviction and a sentence of life in prison.[87]

At least two American companies are attempting to promote and sell fMRI scans as lie detectors. One of these, Cephos Corporation, claims: "Cephos truth verification brain imaging services provides independent validation that you are telling the truth." (See http://www.cephoscorp.com/Cephos_Corp_Home.html) Similarly, No Lie MRI, Inc. claims to offer "unbiased methods for the detection of deception and other information stored in the brain." (See http://www.noliemri.com/) Another company, Brain Fingerprinting Laboratories, uses EEG (electroencephalographic) sensors to record the activity in a subject's brain and to assess whether certain stimuli (e.g. details about a crime allegedly committed by the subject) are present or absent in a subject's memory. (See http://www.brainwavescience.com/ExecutiveSummary.php) The following discussion offers an overview of some of the legal and ethical considerations that have emerged in connection with the expanded uses of brain imaging technology.

Legal Issues

The Use of Brain Images for Deception Detection in the Courts: Before a judge or a jury can consider a piece of scientific evidence, the judge is required to evaluate the admissibility of the proposed evidence – based on specified criteria including scientific validity, reliability and general acceptance in the relevant scientific community. Expert testimony is generally used both to support and to contest the admissibility of novel scientific evidence such as fMRI images. Although commercial developers of lie detection technologies assert that their products are sufficiently reliable for use in court, those courts that have been presented with such evidence have so far declined to rely on it in their rulings. In theory, an accurate test for truth telling could be very useful in assessing the credibility of both defendants and witnesses (and perhaps even prospective jurors); on the other hand, any mechanical test might also be said to undermine the traditional role of the jury as the body for making credibility determinations.

The Use of Brain Images to Assess Criminal Responsibility: In some cases, courts have admitted brain scan evidence to support claims made on behalf of criminal defendants that abnormality of their frontal lobes is a mitigating factor in assessing their responsibility for criminal behavior. More than 30 years ago, in the high profile trial of would be assassin John Hinckley, the court allowed the jury to consider CT scans showing his "abnormal" brain, and the jury ultimately found Hinckley not guilty by reason of insanity. More recently, at least one court has gone so far as to find that a defendant has a right to obtains funding for a brain scan in support of his defense that a traumatic brain injury had caused his criminal behavior.

[87] A. Giridharadas, "India's Use of Brain Scans in Court is Debated." *New York Times* (Sept. 15, 2008)

Existing Legal Regulation of Deception Detection: Existing federal and state laws currently regulate the use of polygraph testing. For example, the federal Employee Polygraph Protection Act of 1988 prohibits employers (with certain specified exceptions) from requiring employees or prospective employees to take lie detector tests. The Act uses a very broad definition of lie detector tests, which arguably can be read to include fMRI-based technologies (which were not in use at the time of the law's passage). The Act's prohibition on lie detector testing applies to "a polygraph, deceptograph, voice stress analyzer, psychological stress evaluator, or any other similar device (whether mechanical or electrical) that is used, or the results of which are used, for the purpose of rendering a diagnostic opinion regarding the honesty or dishonesty of an individual." (Emphasis supplied.) Twenty-five states and the District of Columbia also have laws limiting the use of lie detection devices, some of which are specific to polygraphs and others of which are more general.

In addition, many states have rules regarding the admissibility of lie detector evidence in court proceedings. New Mexico is the only state that allows polygraph evidence to be used as a matter of course in legal proceedings. All other jurisdictions restrict the admissibility of this kind of evidence because of the consensus that its reliability has not been established. In some cases, polygraph evidence can be admitted if the parties to a case agree; in others, for example, in California criminal proceedings, the results of polygraph examinations are explicitly barred.

Proposed Regulation of fMRI-Based Deception Detection: fMRI based lie detectors are not generally subject to FDA regulation, which is limited to oversight of drugs and devices intended for use in diagnosing or curing disease or intended to affect the structure or function of the body. Some commentators have recently suggested that pre-market approval of fMRI deception detection technology should be required by law and could be managed by the FDA or some other federal agency. The purpose of such a requirement would be to ensure that a review process similar to the clinical trial (used for approval of drugs and medical devices) would be used to ensure thorough evaluation of the safety and efficacy of the technology in advance of widespread use.

Constitutional Questions: Under the 5th Amendment of the U.S. Constitution, a person cannot be compelled to incriminate him or herself. However, a person can be compelled by the government to provide physical evidence, such as a blood sample. It is unclear whether fMRI data would be protected by the privilege against self-incrimination or whether it might be classified as physical evidence of patterns of blood flow in the brain, and accorded 5th Amendment protection.

Ethical Issues

The ethical issues associated with fMRI lie detection primarily concern reliability and privacy. Reliability concerns stem from unsubstantiated claims about the accuracy of the results and the resulting risks being imposed on technology users. Privacy concerns arise from the potential for the fMRI technology to gain access to a person's private thoughts without necessarily obtaining permission. These risks of privacy violations are likely to grow over time as MRI functions become embedded in more portable machines that that may allow credibility to be assessed unobtrusively and without the consent of the person whose brain is being examined. These possibilities raise the specter of government intrusion in private matters ("Big Brother") and a general undermining of individual control over others' access to private thoughts.

Additional Selected References[88,89,90,91,92]

[88] D. Meegan, "Neuroimaging Techniques for Memory Detection: Scientific, Ethical and Legal Issues," American Journal of Bioethics: AJOB Neuroscience, Vol. 8, No. 1, 9-20, January 2008. (This journal issue also contains 8 "Open Peer Commentaries" on this article.)

[89] J. Moriarty, "Flickering Admissibility: Neuroimaging Evidence in the U.S. Courts," Behavioral Sciences and the Law, Vol. 26, 29-49, 2008.

[90] J. Rosen, "The Brain on the Stand," New York Times, March 11, 2007.

[91] G. Stix, "Can fMRI Really Tell If You're Lying?" Scientific American, August 13, 2008.

[92] Symposium: Brain Imaging and the Law. American Journal of Law and Medicine, Vol. 33, Nos. 2 & 3, 2007.

APPENDIX E - SURETY METHODS AND TECHNOLOGIES

There are many surety methods and technologies that can support the research and development of cognitive systems. Some more detailed checklists and discussion is presented in this appendix on the following surety areas:

(1) Reliability

(2) Safety

(3) Security

(4) Modeling and Simulation

(5) QMU

(6) Experimental Design

(7) Verification and Validation (framework elements)

Many of the checklists refer to requirements for which design characterization activities may be appropriate or even required in order to provide the necessary evidence that the requirement has been met by the design. Clearly all checklist activities or specific technology discussions may not apply to a specific cognitive system product, particularly during certain product life cycle stages. These “surety methods/technologies” fall in the normative reference category for use within the Evaluation Model, Risk Model, or Maturity Model of the Surety Engineering Cognitive System Framework.

E.1 Design for Reliability

Design for reliability activities can help establish requirements, provide information on the reliability of similar parts, construct reliability models, and use the models to generate preliminary estimates of reliability. Important considerations include knowledge of what has caused failures of similar products, insight into why a design might fail, facilitation of designs that are easy to manufacture and robust in service.

Design characterization activities include preparing fault trees, identifying fault modes and controls for these modes, assisting in the development of sampling plans, and identifying data that will be needed to assess reliability. Reliable designs provide assurance that the manufactured product will be accepted for use and will be available when needed. During deployment, data to assess reliability and support failure investigations is obtained to estimate reliability of existing designs and improve the reliability of future designs. It is critical to have reliability specialists involved as key task members throughout the product life cycle phases.

General:

- Have reliability specialists on the design team.
- Understand the application environments:
- Understand duration of environments.
- Understand cyclical nature of environments.
- Understand combinations of environments.
- Design based on expected range of operating environments.
- Know the reliability requirements. Determine if they make sense and how conformance will be demonstrated.
- Know what has caused failure of similar parts in the past.
- Understand manufacturing processes and the manufacturing environment.

- Identify faults and associated controls and implement the controls during manufacture of development units.
- Prepare an acceptance plan and try it out during development. Determine if necessary data can be obtained.

During concept generation and evaluation:

- Establish requirements.
- Obtain preliminary estimates of reliability.
- Identify environments.
- Identify failure modes.

During design and preparation for manufacture:

- Prepare fault trees and failure mode and effects analyses.
- Assist in the identification of controls for faults.
- Provide estimates of reliability.
- Identify data needed to assess reliability.
- Generate sampling plans.

During deployment:

- Assess reliability.
- Review effectiveness of controls and modify control set as necessary.

Attribute	Potential Attribute Design Characterization Activities
Robust The design is such that there is a demonstrated significant performance margin between the product requirements and the product performance. **Reliable** The design provides for an acceptable probability that the item will perform a required function under stated conditions for a stated period of time.	Design characterization for robustness involves determining limits and margins for various performance parameters. Margin testing can be performed to define early design characterization, characterize development builds, and verify results of process prove in and/or Qualification Evaluation (QE) units. Activities supported by design characterization might include: • Margin testing to establish what the actual margin is or verify whether the required/desired margin can be met by the specified design. • Margins testing to define specified parameters and/or combination of parameters (e.g., shock, vibration, temperature, timing throughput, response time, fracture pressure, and so forth). Requirements specify parameter values, tolerance, and sometimes desirable margins. If such requirements are “fuzzy”, use overtesting to establish the margin bounds and then establish acceptable requirements, tolerances, and margins based on the over testing. Margin testing can be conducted in a variety of ways: • Perform modeling and simulation (maturity of verification and/or validation activities depend on life cycle use) to establish plausible margins and identify where there may be a need to address the margins through actual verification/validation experiments • Consider performing Highly Accelerated Life Tests/Highly Accelerated Stress Screens (HALT/HASS) testing, possibly with specially constructed test vehicles, or perhaps at a higher level of assembly, to identify/verify margin. • Consider performing verification/validation experiments to test robustness in cases where the modeling and simulation, HALT/HASS, or other expert considerations indicate there may be critical concerns • Perform Acceptance Process Validation (APV) testing – Repeat of E- and/or D-Test sequence as evidence of margin for development, PPI, and/or QE builds The margin tests along with production constraints refine the product requirements and provide the basis for production requirements.

E.2 Design for Safety

Safety is required to be a "designed-in" feature of high-consequence products as well as a verified feature of the produced product. Not all components/next assembly/subsystems will have such requirements. It is critical to have safety specialists involved as key task members when there are significant safety product requirements. Safety requirements typically arise in cognitive systems when there is some potential for harm to persons – either from direct invasive actions (e.g., a chip in the brain) or from indirect effects such as might occur from irradiation of the brain area by and external device.

Design characterization might include or support the following activities:

- create and review a product safety theme and allocated safety requirements
- incorporate design features that minimize the possibility of accidental and/or inadvertent safety failures.
- design to meet the numerical requirements through three independent safety subsystems, or obtain deviation based on review, assessment, and concurrence from external technical experts/peer reviews
- design for normal environments to preclude accidental and/or inadvertent safety failures
- design positive measures for engineered safety features that are implemented solely or principally for achieving safety; such measure should be simple, analyzable, testable, repeatable, controllable, provably safe, passively safe, fail safe, fail gracefully (predictably and non-catastrophically), inherently incorporate the safety principles of Isolation, Incompatibility, Inoperability and implement with Independence
- Safety features are primarily identified through Failure Modes, Effects, and Criticality Analysis and Fault Tree Analysis
- conduct electrical characterization to ensure safety-critical lines have adequate margin for isolation, correct signals, peak voltage, positive interruption, bypass preclusion, unique identification of purpose,
- preclude the use of COTS components for safety-critical application features, unless the COTS component design can be verified/validation per first principles
- conduct safety verification/validation/assessment activities throughout concept, development, and production activities
- document a safety assessment conducted by safety assessment experts

E.3 Design for Security

Design characterization for security involves design for access authorization implementation. The project team should include a task member who is expert in security access authentication and control features to be designed and characterized. It is critical to have security specialists involved as key task members when there are passive and/or active security product requirements. Such requirements for cognitive systems may arise from ELS privacy and legal concerns for the operational use of the cognitive system for gathering of individual profile information.

Design Characterization Activities for security protection might include:

- conducting threat analyses to understand how to reduce potential vulnerabilities
- using end to end encryption/decryption technology with modern common authentication module components
- using command disable features to prevent unauthorized use and minimize potential exposure to privacy information

- conduct robust, reliable, and other design characterization activities to ensure that the security features function before, during, and after such activities as specified in environmental scenarios
- conduct repeated testing of development, production, and delivery units to demonstrate product security features

E.4 Modeling and Simulation and Computational Analysis

Limited experimental testing requires high consequence computing for critical design decisions, qualification evidence, and customer acceptance. Significant resources are involved, and systematic processes are key to balance schedule, cost, and computational analysis performance. Flexibility is necessary to provide scientifically defensible calculations. Some of the reasons why a formalized methodology for computational analysis supported by modeling and simulation codes is important to a cognitive system program include:

1. **Support Qualification Where There Is Limited Testing:** One clear reason computational analysis is so important is to support cognitive systems, subsystems, and/or components where the vast amount of information much be analyzed to support the model development. In addition, even component testing using existing facilities such as environmental chambers, EMP platforms, radiation facilities, and mechanical vibration/shock equipment requires availability of the facility, schedule time to conduct the testing, and funding that may not always be possible, at least to the extent desirable. These constraints make it highly desirable to have a cost-effective computational analysis capability to provide verified/validated results when such testing has to be limited.
2. **Provide Design Decision Support Throughout Cognitive System Life Cycle:** Computational analysis can be used throughout the Cognitive System Life Cycle to support design trade-offs, design decision verification, margin computations, and critical parameter sensitivity analysis that would be too costly to perform (or may not be capable of being performed) using available testing methods. In addition, computational analysis can be used to isolate and characterize which experimental tests would be most beneficial to conduct and which design parameters are most critical to measure during the testing process. Depending on the Maturity Level required and capability available, such computational analysis results may also be used for qualification support and/or qualified calculation predictions.
3. **Increase the Confidence of Application Use:** Improved processes and technical activities used in preparation for computational analysis will provide more credible computational results. In turn this results in improved confidence in the computational analysis process, the results from that process, and the repeatability of the process. This computational analysis evidence and confidence in its use is necessary for internal and external review scrutiny.
4. **Improve the Cognitive System Research and Development Effectiveness:** Credible and decisive computational analysis means a more effective balance of cost, schedule, and performance for cognitive system research and development activities: greater confidence in use of computational analysis, reduced costs due to identification of unnecessary testing, more flexibility in meeting schedules, and credible response based on requisite Maturity Level.
5. **Establish Credible Computational Concepts and Adequacy Measures:** Establishment of a standard view of computational analysis concepts and capabilities, along with appropriate adequacy measures, is beneficial and will improve understanding and reduce misconceptions between what computational analysis is required to be and what is capable of being conducted. Needed success criteria will be clarified through appropriate measures whose adequacy is consistent with the requisite Maturity Level and through an improved understanding of the appropriate Maturity Level required to meet a computational analysis requirement.
6. **Facilitate Future Coordination:** Computational analysis activities require coordination among the Computational Analysis Application codes and their deployment, maintenance and support functions, experimental programs, as well as Cognitive System customers. These concepts should

facilitate development of required computational analysis activities and future coordination for these activities.

The fidelity, validation with applicability understanding, uncertainty quantification, and process quality specifically needed depend on the application's computational requirements and the level of risk that is acceptable to the end-use customer, who may be a cognitive system application user or a computational analyst who supports the application user.

There will be some applications where the requirement is to provide key parameter trend information with a potential for wide margin of variance, no required uncertainty quantification, and minimal documented analysis or code development quality processes. On the other hand, there will be some applications where the requirement is to provide state-of-the-art accuracy for computational results along with a clear understanding of what the variances/unknowns/errors are and how they contribute to quantified margins and uncertainty in the results, all traced to a well-documented set of computational analysis, verification, and validation processes and results.

There are four levels of computational analysis formality from low to high that may be required to support Design Characterization activities: Research and Development (R&D), Design Support, Qualification Support, and Qualified Calculation. A description of these levels is provided below and the correlation with the Cognitive System Maturity Model Levels is also provided.

Level 0 – R&D: Computational analysis for research and exploratory projects includes development of mathematical and scientific understanding of physical phenomena that might be applicable to multiple cognitive system mission problems. Also, such activities may be involved with a specific cognitive system application or experiment, but in a purely speculative way, the main purpose being to clarify ideas and concepts.

Level 1 – Design Support: Computational analysis for component design input and decisions includes providing technical input into a design (or other application) problem. The contribution is probably indirect, however, in that other considerations may also used to arrive at a final design decision. Typical for use in prototype development or early full scale development design support.

Level 2 – Qualification Support: Computational analysis to support a component qualification includes analysis that is an indirect contribution to the component's formal qualification or productization. The analysis is used to support qualification, but is not the only source for qualification.

Level 3 – Qualified Calculation: Computational analyses, in particular qualified calculations, are used directly for a component's qualification. The analysis goal for this level is to provide a direct contribution to design and/or production qualification decisions that may or may not be supported by other verification/validation activities. Because the analysis is used to make a qualification decision, the resources used for the analysis must be qualified to support such analysis through a formally defined process. The resources include the analysts, codes used to produce the stockpile computations, and computational infrastructure used to support the analysis.

The Advanced Simulation & Computing (ASC) program at Sandia provides substantive support for design characterization activities. Design characterization support includes:

a. **Multiphysics Codes:** these products are complex, integrated hydro, radiation-hydro, and transport codes for application to design and analysis of experiments, general purpose hydro and radiation-hydro problems, and analyzing radiation and particle transport problems for a variety of applications.

b. **Engineering Codes:** these engineering applications codes support analyses such as thermal and structural dynamics modeling of weapon components and systems under normal, abnormal and hostile environments. Manufacturing process codes support casting, welding and forging operations.

c. P**hysics and Engineering Models:** these products are used for developing science-based models of physical processes, materials response and properties with the goal of improving predictive capabilities. This includes mechanical, chemical and other physical processes.

d. **Material Data Libraries:** these products provide numerically generated representations of material properties and physical data including opacities, cross sections, strength, and other such properties.

e. **Verification & Validation:** includes methods for conducting systematic assessments of predictive capability and uncertainties in primary performance codes and related models to support the needs of stockpile computing. Activities include planning and experiment design, software quality assurance, verification assessments, uncertainty quantification, validation assessments (integral and hierarchical), predictive accuracy estimation and documentation.

f. **Computational Systems:** provides high-computing capabilities necessary to support the ASC program and stockpile computing requirements.

Specific design characterization activities supported by the ASC Program include:

- media radiation solver for characterizing normal and abnormal mechanical environments that arise from impact and accident scenarios (Syrinx)
- capability to model normal and hostile environments with mechanical and vibratory loads (e.g., blast, impulse) and structural response to these loads applied to the exterior of a cognitive system physical structure and propagated to interior components (Salinas/SIERRA)
- intense x-ray environments such as high-intensity radiofrequency fields, pulsed electromagnetic fields such as lighting (EMPHASIS)
- electronic circuit simulation analysis to credibly predict performance under a wide range of operating conditions and environments such as temperature extremes and radiation effects (Xyce)
- numerical modeling capability to simulate high energy density physics environments that solve resistive magnetohydrodynamics equations coupled with thermal conduction, radiation transport, two-temperature ion/electron physics, and an external circuit model (ALEGRA)
- new tools and procedures and integrated them with other tools to assemble a functional capability for quantifying margins and uncertainty (QMU) analysis that has enabled rational assessment of weapon safety in abnormal-thermal environments. An advanced approach to QMU applied to thermal stronglink/weaklink response in weapon systems has been demonstrated. (QMU)
- DAKOTA (Design Analysis Kit for Optimization and Terascale Applications) provides a flexible and extensible interface between iterative systems-analysis capabilities and a broad variety of simulation codes used in the design characterization applications, including structural mechanics, heat transfer, fluid dynamics, shock physics, and many other engineering sciences. DAKOTA provides algorithms for design optimization with gradient-based and nongradient-based methods; uncertainty quantification with sampling, reliability, and polynomial chaos methods; parameter estimation with nonlinear least squares methods; and sensitivity/variance analysis with design of experiments and parameter study capabilities.

E.5 Quantification of Margins and Uncertainty

Quantification of Margins and Uncertainty (QMU) is a decision-support methodology for complex technical decisions centering on performance thresholds and associated margins for engineered systems that are made under conditions of uncertainty. QMU supports management of the nuclear stockpile lifecycle, from driving technical requirements, through design and qualification, to production and maintenance. Thus, QMU can be a valuable methodology to apply to the design characterization activities of a weapon/weapon-related product. Some other views of QMU include:

QMU is the disciplined use of science experiments and testing, modeling and simulation, and expert judgment to ensure adequate performance margins with consideration of known uncertainties in the underlying models and databases.

QMU is a tool used in the Assessment and Certification process for nuclear weapon systems, subsystems and components. QMU is an assessment methodology used to quantify the performance, reliability, safety margins, and associated uncertainties for a weapon in normal, abnormal, and hostile environments. The methodology is centered on a science-based understanding of the physical processes involved, quantification of the likelihood of the dominant failure modes of the system, and the consequences of each failure mode. The primary techniques used in QMU are probabilistic uncertainty analysis, incorporation of experimental testing and surveillance results into modeling and simulation, event tree analysis, and sensitivity analysis.

QMU is the exercise of "due diligence" in the design, qualification and assessment of nuclear weapon systems.

QMU is the disciplined use of testing (lab scale experiments through full system tests when possible), modeling and simulation, and expert judgment to ensure adequate performance margins with consideration of known uncertainties in the underlying models and test databases. Implicit in this approach are sensitivity analyses, the identification (discovery) of failure modes, prioritization of work, peer review, and documentation and archiving of work.

The intersection of QMU and the ASC V&V program is both obvious and critical because the V&V program exists to establish a rigorous foundation of credibility for the computational science and engineering required by the Stockpile Stewardship Program. Without adequate V&V, the analyses performed for QMU will lack demonstrated credibility. The V&V program emphasizes the development and implementation of scientific methodologies that are necessary for the verification and validation of high consequence stockpile computing. V&V is also responsible for establishing the infrastructure and processes necessary for performing non-deterministic analyses (i.e., UQ).

QMU is thus a collection of methods that rest on three key elements, with the goal of supporting nuclear-stockpile decision making under uncertainty. The three key elements of our QMU methodology stress stockpile-lifecycle performance characteristics and are summarized as follows:

- Element 1: Identification and specification of performance threshold(s);
- Element 2: Identification and specification of associated performance margin(s), that is, measure(s) of exceeding performance thresholds; and
- Element 3: Quantified uncertainty in threshold and margin specifications.

QMU quantifies the three major elements (hence, the presence of the word "Quantitative" in QMU) and produces numbers, random variables, or some other more-general measures of uncertainty.

Performance threshold information is typically associated with requirement specifications. There may or may not be tolerances specified as part of the requirement. Depending on the uncertainty, the Margin may or may not be adequate to provide the necessary confidence that the requirement will be met, say with 95% confidence. In general, the Margin is the difference between the required performance of a system and the demonstrated performance of a system, with a positive margin indicating that the expected performance exceeds the required performance. The determination of performance margins is a complex engineering activity that is based on experiment/test, M&S, experience and expert judgment. The consistent application of QMU as part of this information is driven by the demand for demonstrating the science basis for achieving desired margins in system performance.

The quantified uncertainty in threshold and margin specifications is an important link to the decision process as to whether a requirement is met or there is perhaps a requirement gap. It is important to understand that QMU information is just one part of a more comprehensive Risk-Informed Decision Process. There are other inputs to the decision process other than the technical basis that the QMU provides, such as described in SAND2006-5001, Ideas Underlying Quantification of Margins and Uncertainties. Two parts to the uncertainty quantification are important to understand: variability quantification (aleatory uncertainty) and lack-of-knowledge uncertainty (epistemic uncertainty).

When considering QMU, the variability component of uncertainty (U) is expected to be explicitly stochastic, with the needed statistical data underlying its quantification expected to be available. This requires the existence of a statistically significant database. Epistemic uncertainty, and its logical and mathematical distinction from variability in Best Estimate (BE) + U, is very important in stockpile stewardship, where gaps and limitations in predictive capability, incomplete experimental data, and poor statistical databases are common. Epistemic uncertainty is certainly present when there are not enough test data to statistically quantify a presumed aleatory uncertainty.

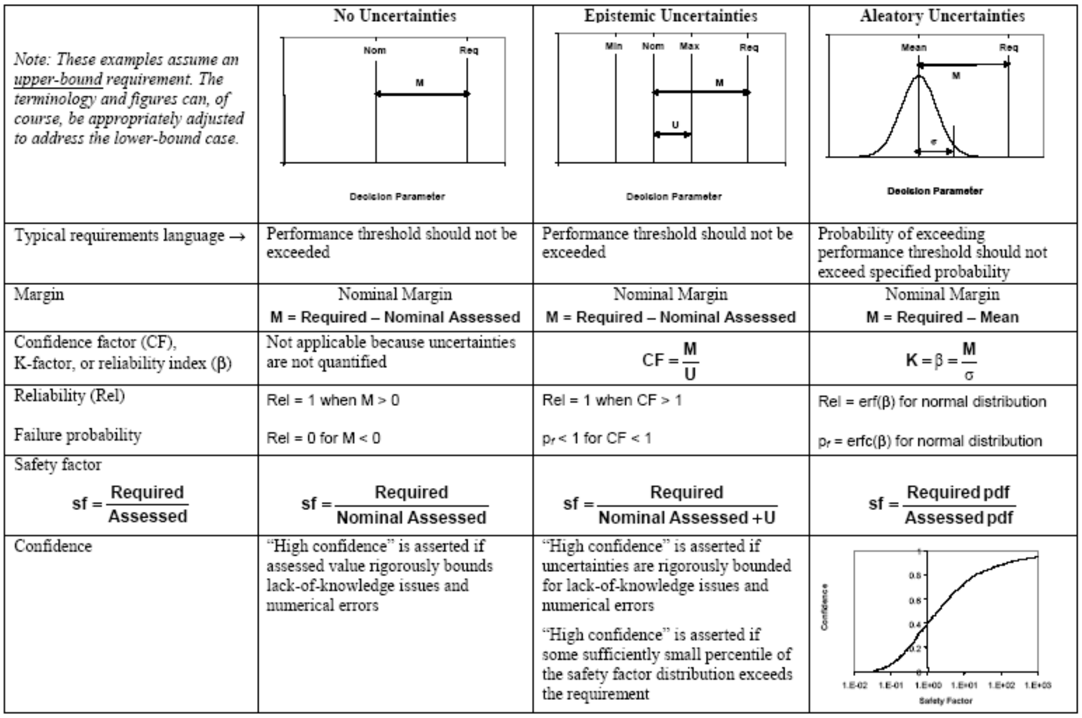
No matter how well we achieve rigorous and credible BE+U, the remaining unknown unknowns, the gaps, and a wide spectrum of constraints force the decision making to still conform to a Risk-Informed Decision Analysis process. Specifically in the stockpile lifecycle, quantified uncertainty U is not a strict substitute for good design principles, use of safety factors, deployment of redundant systems for increased performance reliability, and application of design for computational analysis. It is important to systematically broaden uncertainty ranges beyond what is justifiable in the search for performance cliffs and unanticipated thresholds as well as other decision-threatening regions. Peer review and organizational memory are also critical.

When considering the use of the QMU methodology in a risk-informed decision process (in particular, in support of design characterization risk-informed decisions), the application will be in terms of an underlying scenario associated with the product of interest. Questions of interest are:

1) scenario identification: what can happen?
2) scenario likelihood: how likely is it to happen?
3) scenario consequence: what are the consequences if it does happen?
4) scenario confidence: how much confidence do we have in the answers to the first three questions?

The last question is what involves the use of design characterization activities, including use of QMU. Some elementary terminology in QMU from [SAND2007-6219] is illustrated in Table E-1.

Table E-1. Elementary Terminology in QMU



	No Uncertainties	Epistemic Uncertainties	Aleatory Uncertainties
Note: These examples assume an upper-bound requirement. The terminology and figures can, of course, be appropriately adjusted to address the lower-bound case.			
Typical requirements language →	Performance threshold should not be exceeded	Performance threshold should not be exceeded	Probability of exceeding performance threshold should not exceed specified probability
Margin	Nominal Margin M = Required – Nominal Assessed	Nominal Margin M = Required – Nominal Assessed	Nominal Margin M = Required – Mean
Confidence factor (CF), K-factor, or reliability index (β)	Not applicable because uncertainties are not quantified	$CF = \frac{M}{U}$	$K = \beta = \frac{M}{\sigma}$

Reliability (Rel)
Failure probability
	Rel = 1 when M > 0
	Rel = 0 for M < 0
		Rel = 1 when CF > 1
		p_f < 1 for CF < 1
			Rel = erf(β) for normal distribution
			p_f = erfc(β) for normal distribution

Safety factor $sf = \frac{\text{Required}}{\text{Assessed}}$	$sf = \frac{\text{Required}}{\text{Nominal Assessed}}$	$sf = \frac{\text{Required}}{\text{Nominal Assessed} + U}$	$sf = \frac{\text{Required pdf}}{\text{Assessed pdf}}$
Confidence	"High confidence" is asserted if assessed value rigorously bounds lack-of-knowledge issues and numerical errors	"High confidence" is asserted if uncertainties are rigorously bounded for lack-of-knowledge issues and numerical errors "High confidence" is asserted if some sufficiently small percentile of the safety factor distribution exceeds the requirement	

E.6 Experimental Design

Experimental Design (ED) enables engineers to study the effects of several variables affecting the response or output of a certain experimental process. ED methods have wide potential application in the engineering design and development stages. It is the strategy of weapon engineers to develop products and processes insensitive to various sources of variation using ED. The potential applications of ED include:

- reducing product and process design and development time;
- studying the behavior of a process over a wide range of operating conditions;
- minimizing the effect of variations in manufacturing conditions;
- understanding the process under study and thereby improving its performance;
- increasing process productivity by reducing scrap, rework, cost of quality
- improving the process yield and stability of an on-going manufacturing process;
- making products insensitive to environmental variations such as relative humidity, vibration, shock and so on; and,
- studying the relationship between a set of independent process variables (i.e., process parameters) and the output (i.e., response).

The following steps illustrate what experimental design might require:

1. Definition of the objective of the experiment.

2. Selection of the response or output.
3. Selection of the process variables or design parameters (control factors), noise factors and the interactions among the process variables of interest. Noise factors are those which cannot be controlled during actual production conditions, but may have strong influence on the response variability. The purpose of an experimenter is to reduce the effect of these undesirable noise factors by determining the best factor level combinations of the control factors or design parameters. For example, in an injection molding process, humidity and ambient temperature are typical noise factors.
4. Determination of factor levels and range of factor settings.
5. Choice of appropriate experimental design.
6. Experimental planning.
7. Experimental execution.
8. Experimental data analysis and interpretation.
9. Experimental documentation

Probably the most well-known method (normative reference) for conducting experimental design is called Design of Experiments (DOE). DOE refers to an experiment where one or more variables believed to have an effect on an experimental outcome are identified and manipulated according to a plan. The key elements of the experiment are:

1. Response variable: The outcome variable being investigated. Also called independent variable
2. Primary variables: The controlled variables believed most likely to have an effect on the response variable. Also called independent variables.
3. Background variables: Variables, identified by the designers of the experiment, which may have an effect but cannot or will not be deliberately manipulated or held constant.
4. Common causes or experimental error: those variables not considered explicitly in the experiment. This is the "noise" measurement for the experiment.
5. Interaction: A condition where the effect of one factor depends on the level of another factor.

All effects are statistically compared to the noise, and essentially a "signal to noise ratio" is calculated, called the "F-ratio". Large F-ratios are deemed to be significant in the statistical sense. Statistical significance does not necessarily mean "important" in a business or engineering sense. The design characteristics of DOE include:

1. Replication: The collection of more than one observation for the same set of experimental conditions.
2. Randomization: Applying the various experimental treatments in an order that is without any pattern.

It is generally recommended that random numbers be used to determine the order of applying various treatments. There are various experimental models that might be selected depending on the application:

1. Fixed effects model: An experimental model where all possible factor levels are studied.
2. Random effects model: An experimental model where the levels of factors evaluated by the experiment represent a sample of all possible levels.
3. Mixed model: An experimental model with both fixed and random effects.
4. Completely randomized design: An experimental plan where the order in which the experiment is performed is completely random.

5. Randomized block design: An experimental design where the experimental observations are divided into "blocks" according to some criterion. The blocks are filled sequentially, but the order within the block is filled randomly.
6. Balanced design: A randomized block design where each combination appears the same number of times.
7. Unbalanced design: A randomized block design where different combinations appear a different number of times.
8. Latin-square designs: Designs where each treatment appears exactly once in each row and column; Useful when you want to allow for two sources of variation; A third variable is associated with the rows and columns in a carefully defined fashion; the number of rows, columns, and treatments must all be the same; no interactions between the row and column factors.

Analysis of Variance, Factorial Experiments and other methods also complement the Experimental Design approach.

E.7 Verification and Validation

In SAND2008-6453[93], the concept of applying verification and validation as normally applied to a physical science-based domain such as weapons is extended. A framework is proposed that extends the principles of the Advanced Scientific Computing (ASC) approach primarily applied to weapon system applications into the area of computational social and cognitive modeling and simulation. This V&V framework is an essential element of the Surety Engineering Cognitive Systems Framework Evaluation Model for application instances where the cognitive system is represented by a computational system implementation. The framework concepts are also applicable for V&V of any model instances (Specification, Design, Evaluation) for a cognitive system instance of the Surety Engineering Cognitive Systems Framework.

The SAND2008-6453 report argues to 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. This is precisely the approach needed to integrate system/surety engineering with ELS engineering within a V&V umbrella for model evaluation. As mentioned in the report, 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 can be done only with the presumption of adequate model maturity – or at least adequate evidence of the maturity of the modeling and simulation tools as well as the real-world cognitive system representations being modeled.

Many of the methods described in the above sections of this Appendix are described in the referenced report. In particular, a clear distinction is made between the terminology for "verification" and "validation":

> *Verification* is the process of determining that a computational software implementation correctly represents a model of a physical process

[93] SAND2008-6453, "R&D for Computational Cognitive and Social Models: Foundations for Model Evaluation through Verification and Validation," Laura A. McNamara, Timothy G. Trucano, George A. Backus, Scott A. Mitchell, Alexander Slepoy, September 2008.

> 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.

This report provides the necessary framework for implementing many aspects of the Evaluation Model – both from a normative reference as well as specific V&V activities necessary to provide varying levels of confidence in the cognitive system implementation. In particular, the specific application of V&V methods within the cognitive system domain is investigated, and although primarily targeted to computational implementations, can be applied more generally to implementations that may be biological and/or hybrid.

The following Figure E-1 from the report is an illustration of application of ASC V&V tools and approaches to Cognitive/Social Modeling and Simulation V&V.

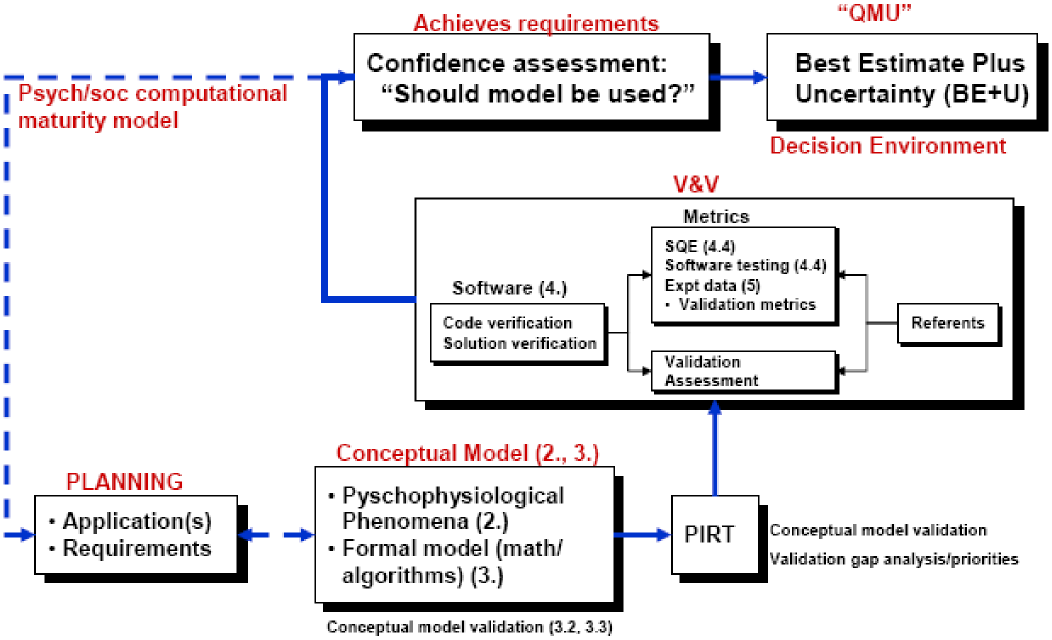

Achieves requirements
Confidence assessment: "Should model be used?"
"QMU"
Best Estimate Plus Uncertainty (BE+U)
Decision Environment
Psych/soc computational maturity model
V&V
Metrics
SQE (4.4)
Software testing (4.4)
Expt data (5)
• Validation metrics
Software (4.)
Code verification
Solution verification
Referents
Validation Assessment
PLANNING
• Application(s)
• Requirements
Conceptual Model (2., 3.)
• Pyschophysiological Phenomena (2.)
• Formal model (math/ algorithms) (3.)
Conceptual model validation (3.2, 3.3)
PIRT
Conceptual model validation
Validation gap analysis/priorities

Figure E-1. Cognitive/Social Modeling and Simulation V&V

This report provides process steps for conducting V&V and illustrates in the report Appendix B the V&V approach for cognitive systems within the context of the memory reasoning derivation presented in this report in Appendix F.2.

Numerous references are also included in the SAND2008-6453 report that provide depth to the Evaluation Model of the Surety Engineering Cognitive System Framework. The original reference on the Predictive Capability Maturity Model (PCMM) for Computational Modeling and Simulation[94] provides a maturity model for the verification and validation related to science-based modeling and simulation. For cognitive system, the PCMM complements the overall surety engineering framework CSMM maturity model. Recent internal reports[95,96] have significantly enhanced the PCMM details.

[94] W. Oberkampf, M. Pilch, T. Trucano, "Predictive Capability Maturity Model (PCMM) for Computational Modeling and Simulation, " SAND2007-5948, October 2007.
[95] M. Pilch, "PCMM 2nd Generation," internal SNL Word document, September 30, 2008.

APPENDIX F - EDUCING INFORMATION RESEARCH

F.1 Educing Information Concepts

From the Educing Information Report[97], the use of the term “educing” is coined to reduce negative perceptions that have grown out of the use of the term “interrogation”. The study team acknowledges that this term has not yet come into widespread use, but more accurately describes the scope of this important study.

The definition of Educing Information (EI) encompasses:

(1) “elicitation” - engaging with a source in such a manner that he or she reveals information without being aware of giving away anything of value),

(2) “strategic debriefing” - systematically covering topics and areas with a voluntary source who consents to a formal interview, and

(3) “interrogation” - interaction and conversation with a source who appears initially unwilling to provide information.

> “EI implies a “system” of gathering information about and from a source and a spectrum of approaches, tools, activities, and techniques. This may involve investigative efforts, development of scenarios, and involvement of others (teams of interviewers and analysts, willing sources, and collaborators). Effective practice of EI usually extends beyond one-to-one interactions with a source.”

The basic problem addressed by this report is that “the scholarly and scientific community has not systematically studied eduction for 45 years.” This is at the heart of what the Surety Engineering Framework for Cognitive Systems outlined in this SAND report is addressing – the use of a systematic framework within which the cognitive science technologies that could very well support EI can be studied and their maturity and value identified.

The concepts of ELS directly apply to the EI activities in that the process is required to follow the “rule of law” and “encouraged” to apply two principles: (1) that an individual is innocent until proven guilty, and (2) that civil rights are inherent in Constitutional Law (at least the US version).

The goal of the interrogation part of EI is to develop the truth. Although interrogation activities have been conducted for many centuries, the consensus of research evidence points to the lack of empirical studies in the social and behavioral sciences to directly address the effectiveness of interrogation in general practice, or of specific techniques. The “effectiveness” of existing interrogation techniques has been accepted without sufficient scrutiny.

The three barriers to EI success in the current generation of practices are identified as: (1) linguistic and cultural, (2) scientific and technical, and (3) interpersonal and intrapersonal. This suggests that not only should an evaluation model include scientific and technical improvements, but one must consider ELS issues that typically address the personal, language, and cultural aspects of cognitive system technologies.

[96] M. Pilch, T. Trucano, “PCMM Layout,” internal SNL PowerPoint presentation, October 27, 2008.

[97] “Educing Information Interrogation: Science and Art, Foundations for the Future,” Intelligence Science Board Phase 1 Report, National Defense Intelligence College, December 2006.

The report study team believes that EI is envisioned as a process rather than as an applied set of techniques. This vision fits well with the process-oriented aspects of the Surety Engineering Framework. The integration of multiple fields of discipline is viewed as critical to success. Such fields include many areas of cognitive system interest such as psychology, linguistics, communication as well as the empirical research and development methods attributed to engineering and science-based research. Four specific major conceptual models (and associated theories) are identified that would fit the Surety Engineering Framework Design Model concepts and to some extent the associated normative references: communications, discourse analysis, persuasive message production/analysis, and negotiation theory. The report provides extensive discussion and references for the major conceptual models. The key aspect is to integrate the multiple disciplines of cognitive systems research, social science literature research, and scientific engineering and surety methods. In particular, two "new age" technologies associated with cognitive systems were described as having attracted interest: neurolinguistic programming and subliminal persuasion. To date the statement is made that existing research has not supported the claims of either of these technologies – at least in regard to such technologies affecting internal mental processing to support an effective strategy for EI.

In regard to cognitive systems and the potential for supporting EI, the report indicates that educing information from a human source requires some understanding of how people in general acquire, process, store, and retrieve information. Without this knowledge, it is possible to misinterpret or even contaminate stored information so that not even the source can any longer discern the "real truth." This position is clearly represented in the Surety Engineering Framework Design Model. The report describes the normative reference "modal model" of memory (sensory memory, short-term memory, long-tem memory) is clearly part of the research represented by the Design Model. In addition, environmental effects such as stress, fatigue, distraction, duress, pain, and sensory deprivation are acknowledged to affect not only the source's motivation to provide accurate, useful information, but also the capacity to do so. So, the Surety Engineering Framework Specification Model requirements for behavior, structure, and environment are all represented as concerns for would appropriate evaluation methods and operational scenarios must be derived.

One of the report sections addresses the research on detection deception as an area of EI. Two primary approaches to detecting deception rely on psychophysiological (e.g., physical responses) and behavioral (e.g., actions, statements or responses monitored through observation) indicators. These approaches are clearly within cognitive system applications. In particular, the lie-detector aspects of the polygraph and fMRI are particularly applicable. Overall, extensive data to this point suggest that for all groups, novice to professional, accuracy in determining when someone is being deceptive is only marginally better than chance. This conclusion supports the arguments discussed in Section D.2.7 of this SAND report concerning the fidelity of the fMRI for EI use and the associated ELS issues that result. The report describes alternative methods that are emerging and merit study, for example, probability theory and covert physiological sensor systems. This is just a limited view of the potential use of the Surety Engineering Framework – particularly in the Evaluation Model and its use of techniques associated with reliability, safety, security, QMU, probabilistic methods, Modeling and Simulation, experimental design, and verification and validation. Probabilistic sampling theory can assist in deriving more accurate conclusions concerning experiment populations that frequently bias the experimental evidence against real-world populations of interest. These deficiencies also elicit ELS concerns as described in Appendix D - Cognitive Systems and ELS Risk Research.

F.1.1 Description of Detection Deception Applications

The Educing Information report provides some interesting descriptions of mechanical detection deception technologies. The general conclusion is that there needs to be a layering of technologies to provide more assurance that Type I and Type 2 errors are reduced. This "layering' concept is very much like the safety and reliability technologies for reducing the potential for failures. In the case of safety techniques, multiple regions of protection are created that have to be penetrated through different failure modes in order for safety to be compromised. With each layer there is an associated probability failure distributions associated with threat scenarios. In a similar way, there might be several layers of detection deception mechanisms with associated probability of failure distributions associated with threat scenarios. The failure distributions would be associated with the source population and the ability of that source population to defeat the detection deception mechanism – either purposely or not. Research would be required to verify and validate any conclusions from this integrated approach to detection deception. The Surety Engineering Framework for Cognitive Systems would provide a systematic mechanism within which such research might be conducted.

Some of the specific psychophysiological detection deception mechanisms include:

(1) polygraph with Guilty Knowledge Test (GKT) and Concealed Information Test (CIT)

(2) Electroencephalography (EEG)

(3) Radar Vital Signs Monitor (RVSM)

(4) facial expressions, eye blinks, saccades, fixations

(5) Voice Stress Analysis (VSA) and Computer Voice Stress Analysis (CVSA)

(6) thermal imaging

(7) truth serums/narcoanalyis

(8) Laser Doppler Vibrometry (LDV)

(9) Eye Movement Memory Assessment (EMMA).

Research concerns focus around the lack of reliability, use in unintended scenarios, invasiveness, portability for field use, cost to develop/support/operate, and timely availability of the technology. The bottom line is that individually, none of these mechanisms is viable enough to provide much confidence in supporting deception detection. To some extent research has indicated the use of multiple techniques provides improved support.

Dissatisfaction with the lack of a clear causal chain from the psychological decision to deceive, to the autonomic functions (e.g., skin conductance, respiration) currently measured by the polygraph, has led some researchers to seek measurements that are closer to the biophysical seat of decision making. The field of neuroscience has long sought to "understand the biological basis of consciousness and the mental processes by which we perceive, act, learn, and remember" (Kandel, 2000, p. 5). Revolutionary improvements in neuroscientific techniques, combined with the sophisticated signal processing techniques made practical by advances in information processing technology over the past few decades, have made it possible to observe the neurophysiological processes of the brain itself with increasingly greater resolution in time and space. Some of the advanced techniques for studying the relationship between cognitive and neural processes include:

(1) Electroencephalography (EEG)

(2) Magnetoencephalography (MEG

(3) Positron Emission Tomography (PET)

(4) functional Magnetic Resonance Imaging (fMRI)

(5) Near Infrared Spectroscopy (NIRS)

(6) Transcranial Magnetic Stimulation (TMS).

The interesting perspectives on the potential of these technologies provides some insight into the issues that must be addressed by the Surety Engineering Framework for Cognitive Systems.

(1) Invasiveness of some techniques creates safety, privacy, reliability and ELS concerns

(2) Effectiveness of countermeasures (intended and/or unintended)

(3) Cost of equipment and technical expertise, and lack of portability

(4) Maturity of the technology to cover the application domain of interest

The most significant problem is that none of these mechanical devices has been scientifically shown to be capable of accurately and reliably detecting deception. In particular, none of the technologies has been proven to be any better than the basic polygraph. The report indicates there are two schools of thought on how to address the problem: theory first and system first.

Those who subscribe to the “theory first” school of thought believe that additional research is needed to assert and test hypotheses that explain why lying causes measurable changes (somatic, autonomic, or neurological), and not simply to establish a correlation between the act of lying and particular values of, or changes in, the observed features.

Those who subscribe to the “system first” school of thought believe that it is possible to develop a functional and useful system without waiting for the development of an underlying theory that is universally accepted by the scientific community.

Unfortunately, aspects of both of these “schools” are needed. Without theory, we are driven to an exhaustive testing of all possible combinations of factors – clearly impractical. The likelihood of arriving at even a partial theory much less a comprehensive one in the near term is clearly low – particularly due to the large uncertainty and lack of maturity of the existing technologies. What is not stated in the report is the need for a systematic approach such as the Surety Engineering Framework within which to study the maturity, risk, reliability, safety, security/privacy, and evaluation fidelity of such technologies – so we know what we don’t know as well as what we do know. Particularly in the cognitive system applications, there is such a large unknown at this time that any progress toward problem solutions must incorporate such a systematic approach to representing the problem domain, design solution, and evaluation results.

F.1.2 Description of General Surety Application Strategies

The bottom line for the importance of Educing Information is that the very identity of our nation and its national security hang in the balance. From being able to simply determine the truth of information provided by compliant sources to educing whether information from or about potential terrorists is accurate – we need to understand what we know and what we do not know. We need to have confidence in the measurements, the variance of the measurements, and the uncertainty in those measurements. Since EI is so closely linked to cognitive information and processing, cognitive systems will be a potential area of support – but, there needs to be a systematic approach to knowing whether a given system, component, technology or combination is reliable, safe, secure, and satisfies our innate interpretation of the "rule of law" and its associated ELS issues.

The Surety Engineering Framework for Cognitive Systems provides a necessary but not sufficient set of models that help define what needs to be done and how it needs to be represented. Normative references provide the known theory/practice. Models of Specification, Design, and Evaluation provide the basic engineering representation. The Maturity and Risk models provide the mechanism for capturing and representing vulnerabilities and threats in the use of cognitive systems for support of EI goals. In short, the Framework provides an approach to systematically defining what we do know and what we do not know about a given cognitive system/technology. Basic principles of surety and ELS continually guide the integration of potential cognitive system concepts/technologies and act as a check and balance on the scientific evidence necessary to state convincing claims and arguments.

Yonas and Jung in their short article[98] describe the importance of using a Neuroscience Engineering (NE) approach. The NE approach "is designed to focus on problem-solving based on iterating theory, experiments and modeling to satisfy ultimately the needs of people with real applications and products." This is precisely the focus of the Surety Engineering Framework for Cognitive Systems. Yonas and Jung continue to describe this approach as:

> "…extremely multidisciplinary involving neuroscientists, psychologists, physicians, engineers, sensor and information system developers, modeling and simulation using high-performance computing and, of course, ethicists, as we are challenged with the prospect of pushing human capabilities to new heights. Since the people in these various disciplines all have their own languages, definitions, and even axes to grind, just bridging the gaps is a major undertaking, and connecting these complex disciplines with the creation of practical systems solutions is … a challenge."

In relation to the detection deception aspect of Educing Information, Yonas and Jung further believe that:

> "Another potential application of NE is in the field of deception detection, a field that has mostly relied upon the polygraph, a 70-year-old idea that basically uses several indicators of stress, and which has been widely denounced as having no scientific basis. With advances in brain functional measurements, researchers should be able to better understand what parts of the brain are calling the shots during deception and under what circumstances."

[98] G. Yonas, R. Jung, "Sandia and Mind Research Network Have Got a New Discipline," Innovation: America's Journal of Technology Commercialization June/July 2008.

F.2 Derived Cognitive System Instance: Human Memory and Reasoning

One Sandia project on "Modeling Aspects of Human Memory and Reasoning" [99] illustrates a cognitive system research project where the specification model, design model, evaluation model, risk model, and maturity model represent a particular instance within the cognitive system surety engineering framework. The sections below describe some aspects of the model fragments to illustrate the application. This is particularly useful in that the prime area of application focus is on a higher fidelity Design Model instance and a computational implementation of part of that instance. The application Design Model is derived from the framework conceptual Design Model. In addition, some attention has been taken to describe research requirements and some evidence activities involving verification and validation. Some of the project descriptions are reworded somewhat to fit more of a requirements and evidence description. There is still room for improvement in addressing surety-specific aspects as well as potential ELS concerns. Some of these potential "gaps" are identified, but not analyzed in detail.

F.2.1 Case Study Background and Normative References

This Sandia research effort is extending the current Sandia Cognitive Framework (conceptual design model, see Section 3.3) by incorporating a representation of memory processing, focusing on hippocampus and neocortical systems described in current complimentary learning systems theory (i.e., normative reference: cortical-hippocampal theory of declarative memory, Eichenbaum[100]). This design model extension will also specify how hippocampal and cortical representations interact at multiple levels of abstraction to support the interleaving of new information within the cerebral cortex. For example, the perceptual features of relational memory processing are being integrated into an existing Sandia computational model. This project is intended to produce a neuro-cognitive computational architecture that represents episodic memory. The FY08 work greatly extended current computational models (e.g., the normative reference: McClelland[101]) wherein a pre-existing knowledge structure in cortical areas is challenged to incorporate new information within an existing network. To accomplish this goal an innovative neurocognitive representation is produced that leverages the surety engineering cognitive system design model that is a preexisting cognitive architecture developed at Sandia.

Research has found that the hippocampus plays a central role in forming and temporarily storing representations of personal experiences. These representations are later migrated to widespread areas of the cerebral cortex, which are then permanently stored. Existing work has produced a computational model that represents the fundamental features of hippocampus-dependent relational processing. Current efforts plan to test this representation against human memory by comparing the performance of normal human subjects and people lacking normal hippocampal function with the performance of the full model and the model without a functional "hippocampus" on the same memory tasks. Success of the model will be measured in terms of similar performance compared to that of humans with and without the contribution of hippocampal processing, as a function of experimental parameters and controls. The basic project concept is illustrated in the following key bullets:

[99] Bernard, Michael et al. "Memory & Reasoning LDRD Design & Testing Report," SAND2008-xxxx, 2008.

[100] H. Eichenbaum, "Hippocampus: Cognitive processes and neural representations that underlie declarative memory," Neuron 44 pp 109-120, 2007.

[101] J. McClelland, B. McNaughton, R. O'Reilly, "Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory," Psychological Review, 102(3), pp 419-457, 1995.

- **Model purpose:** *social interaction realism to mindset forecasting*
- **Information:** *neuro-psycho-socio theory/research & report data*
- **Scalability (neuro →← cultural):** *neuro-psychological*
- **Optimal complexity (realism →← reductionism):** *realism*
- **Level of predictability (model/domain):** *medium*
- **Metrics:** *Human subject experiments*
- **Models:** *neuro-cognitive network model of humans*
- **System utilization:** *PDA to distributed supercomputer*

F.2.2 Case Study Specification Model

To date, this project has been focused on project-level requirements that translate loosely into what might eventually evolve into a product-level requirement specification. A Design Document for Specific Psychophysiological Functionality has been constructed to guide the project requirements. Some of the normative references that are the basis for this research include the cortical-hippocampal theory of declarative memory, Eichenbaum and McClelland. These normative references are illustrative of the integrated representation of requirements and design/architecture models from literature – rather than a systems engineering product view of design/architecture being a derived representation of an application requirements specification. This is typical of the maturity level for research areas.

For this particular case study example, the following “requirements” represent the “specification model” instance.

1. Psychophysiological Phenomenon Requirements

Shall provide an extensive review of the psychological phenomenon to be modeled and simulated.

PP.1 Shall Review of Theories, Research, & Psychophysiological Models

Provide a general review of the literature, including a review of past and current theories and research. This should be akin to an Introduction section of a review paper.

Framework Note: This also identifies the specification of the normative references for the Specification Model from which to determine potential “gaps” in existing approaches. The fact that these “gaps” may be advances in the theory is what may need to be “proved”.

PP.2 Shall Provide Theoretical Basis of Phenomenon

Provide a more specific discussion of the selected theory and its supporting research. Justify why this theoretical model was selected over other theoretical models.

Framework Note: This provides a justification for reducing potential “gaps” between the normative references and the selected theoretical model. This also establishes the basis for potential fidelity measurements relative to verification and validation evidence provided by the research results.

PP.3 Shall Provide 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?

Framework Note: This provides the context for the environmental fidelity, both addressed by the research project and not expected to be addressed by the research project. This could also identify further research to close credible environment scenarios that have not been addressed prior to detailed prototyping and/or productization.

PP.4 Shall Provide 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.

Framework Note: The feasibility claim/argument provides substantive evidence of the model fidelity (and non-fidelity) relative to the various characteristics identified as "Embodiment".

2. Formal Model Requirements

FM.1 Shall formally express the psychophysiological model.

Note: It is anticipated that the team will need to develop this model from or building upon multiple sources in the literature (normative references for the model), 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.

Framework Note: The formal model requirements also provide the basis for the model fidelity and any potential "gap" the result from the implementation in comparison with the theoretical basis.

FM.2 Shall Provide Psychophysiological Model of Phenomenon

This is the principal section that specifies the model. It should be written in clear prose that is accessible to all readers with relevant specialties (particularly psychologists and computer scientists) while expressing the model with enough formality that there is no ambiguity on how its inputs map to its outputs. A possible template for the specification might be…

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.

Framework Note: The formal model phenomenon requirements primarily from an environment scenario perspective. This provides the contextual basis for the environments in which the model may or may not have fidelity and any potential “gap” the result from the implementation in comparison with the theoretical basis.

3. Verification and Validation Requirements

VV.1 Shall Validate Embodiable Model (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.

Framework Note: Although the project has identified this requirement as “if needed”, it is highly likely that independent of whether the model is adapted or integrated that the validation evidence will still need to be either identified and “re-validated” or identified and “validated”. In other words, validation will be required. Due to project constraints, the validation may not be able to be conducted as part of the project, but will be required prior substantive use of the project’s research results. How well this is done in this project will determine the verification and validation fidelity and to a great extent any gaps in surety engineering aspects of V&V that have not been conducted. This may point back to the model and its computational implementation (design architecture as well as software code implementation) for identified gaps. These gaps may not be a concern for a research effort, but might be for more critical decisions or productization of the processes/products from this research. Also, this validation activity should probably be considered to be a part of the "Full Experimental Validation" activity – captured simply as Validation. Another possibility is that this validation activity may well be a verification rather than validation activity.

VV.2 Shall Verify Psychophysiological Model

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.

Framework Note: This is a similar requirement to validation. In this case, there is a somewhat better process description of what is required of the verification activity. The completeness of that as part of the Evaluation Model (normative references for standard methods as well as actual evaluation activity implementation) will determine the fidelity and maturity of the verification part of the system/surety engineering framework.

VV.3 Peer 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.

Framework Note: This is actually part of the verification activities as a best practice – depending on just how the "independent" review is actually conducted. The requirement should be without caveats – that is, it does not matter who developed the model, it should be peer-reviewed by an "independent" review team. The level of reviews may go from internal to SNL, but part of project team, internal to SNL, but external to project team, and external to SNL with various types of participants: subject matter experts, public focus groups, and so forth depending on the project results and their potential impact.

4. Software Implementation of Model Requirements

This set of requirements is to ensure the software implementation can execute/simulate the model with some level of fidelity.

SW.1 Shall Implement 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.

Framework Note: The software requirements involve defining numerical algorithm solutions to the mathematical algorithms of the psychophysiological model and defining any implementation dependencies (computer language, computational infrastructure such as numerical accuracy/error bounds). Although the requirement notes indicates "arbitrary" computer language – the intent of this requirement needs further explanation. A pseudo-code specification of the numerical algorithm would be one possible translation, but there are still likely to be computational implementation dependencies that would need to be specified. This might limit the potential software implementation languages as well as computational infrastructure (e.g., specific computers, data analysis, visual representation of results). These potential dependencies are required to be specified. Note that the stated requirement should have specified Psychophysiological rather than just Psychological.

SW.2 Shall Provide 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.

Framework Note: This “account” is really the basis for a design specification that accompanies the Design Model instance information. Although there is certainly value in “prior to actual code-writing”, there is also value in prototyping code solutions to understand what can and can not be done – and to then refine and update the actual software requirements. Note that the Algorithm translation information of SW.1 is actually part of the software requirement specification to which this software design specification corresponds.

SW.3 Shall Document Software Implementation

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).

Framework Note: “literate programming practices” isn’t quite a normative reference – there are normative references that exist for best practices – including coding practices. Trying to summarize in a paragraph such information is probably not the best way. This probably will create a “gap” between the actual coding practices used and what a normative reference might require. The implementation of the software (via software inspections/formal reviews) would be a best practice to check the practices. Standardized tool sets such as Doxygen can also set the standards and check them. This type of information is termed the “as-built” detailed design information – and does not necessarily take the place of a well-specified design architecture document.

SW.4 Shall Verify the Code Implementation

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.

Framework Note: The medium for documenting the “Software model” would be a design architecture specification. That would represent to the developers of the algorithmic models how the numerical methods satisfy the algorithms and are mapped into the top level software design architecture. The other review mechanisms would be actual code reviews to ensure that the numerical algorithms have been adequately implemented in the code. Use of the Formal In-Process Review (software inspection) would be the desired normative reference for the review process. This process could be used for both the phenomenon model to software model verification and the software model to software code implementation verification.

5. Full Experimental Validation Requirements

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.

XV.1 Shall Develop 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.

Framework Note: This is a good objective of this research effort and can address much of the "validation" fidelity – whether conducted during the research, next stage prototyping, or some later productization effort. To address it early is the correct approach in accordance with the Maturity concept and "gap" analysis of the framework.

F.2.3 Case Study Design Model

The following figures illustrate the relationship of the project instance Design Model (actually several levels of design model representations) and how it relates to the conceptual Design Model in the framework. This work was developed under the SHERCA (Sandia Human Embodiment and Representation Cognitive Architecture) project prior to 2007.

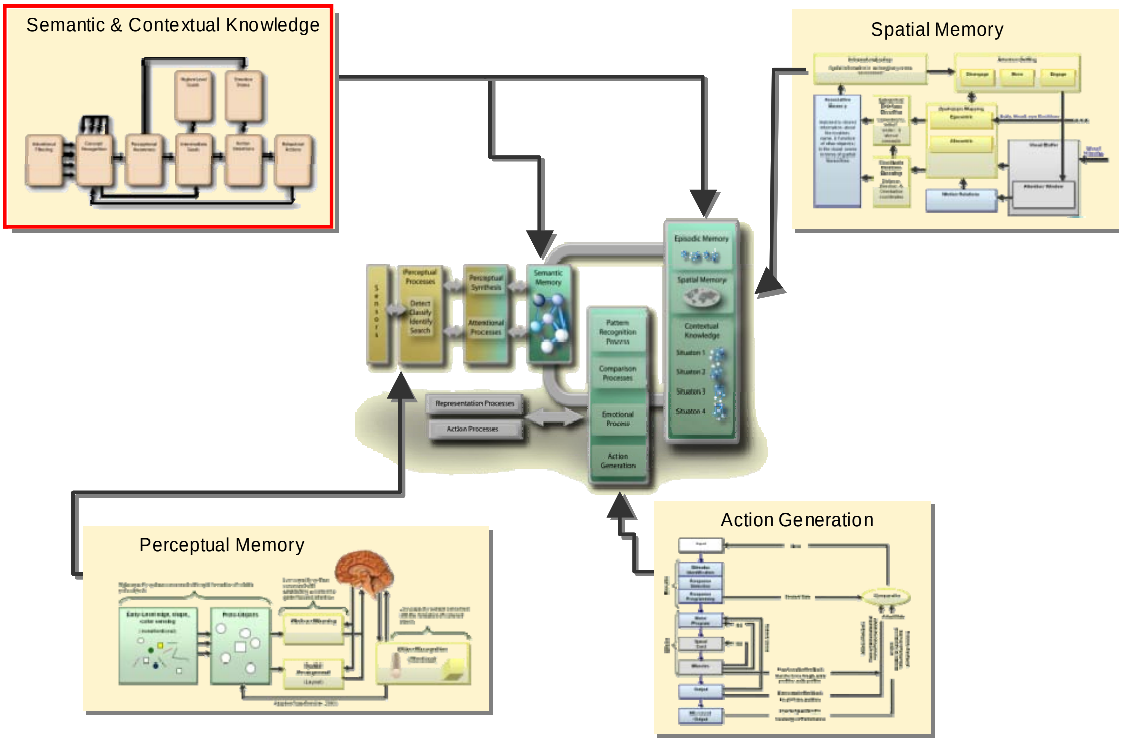


Figure F-1. Neuro-Cognitive Episodic Memory Design Model

The upper left corner of the model captures the episodic memory representation. This part of the design model is illustrated in more depth in Figure F-2.

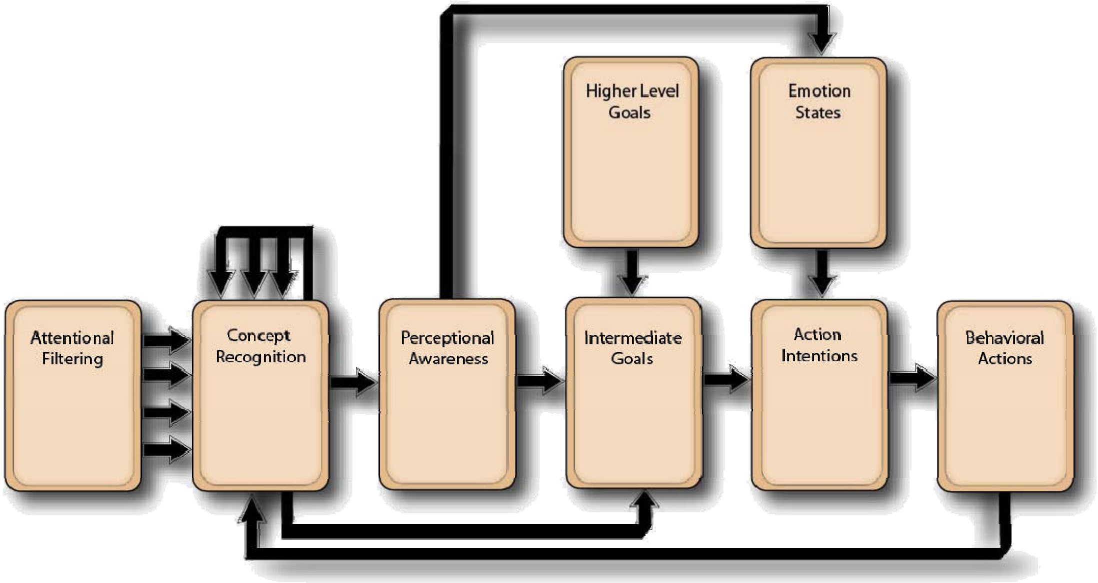


Figure F-2. Simulated Cognitive Functioning Design Model

This Cognitive Process Diagram displays the activation flow of perceptions, concepts, states, and actions. Many of these cognitive processes also affect specific cultural and emotional behaviors. Briefly, the primary processing functions of this Design Model include:

Attentional Filtering: The ability to perceive concepts/stimuli (objects or humans) if in field-of-view and attended. Simulated humans can determine if another simulated human is looking at them or other objects or humans.

Concept Recognition: This top-down “cognitive” activity allows certain concepts/stimuli to prime other concepts/stimuli through an activation network.

Perceptional Awareness: Perceptual awareness gives the ability to recognize a given context by recognizing certain patterns of related concepts or environmental stimuli.

Goals: Higher Level and Intermediate: Agent has a hierarchical set of high-level goals (e.g., protect family), and intermediate goals that are activated to carry out higher-level goals. The intermediate goals that are fully activated are determined by the agent’s perceptions and the stimuli that are present in the environment.

Emotion States: Emotions (degrees of fear and anger) are dynamically mediated or provoked by perceptions that simulated human might have at any given moment.

Action Intentions: The combined activated perceptions and intermediate goals create an action intention. Anticipated actions are mediated by the dynamic emotion states and the Intermediate goal states of the simulated human.

Behavioral Actions: The behavioral action that is selected is a function of the current emotion state and the intended action. High-level actions are chosen and performed through selections via the cognitive model.

The Sandia computational simulation of the human cognitive architecture that implements the Design Model as illustrated in Figure F-2 is illustrated in Figure F-3. This is part of the Design Model from the software implementation viewpoint.

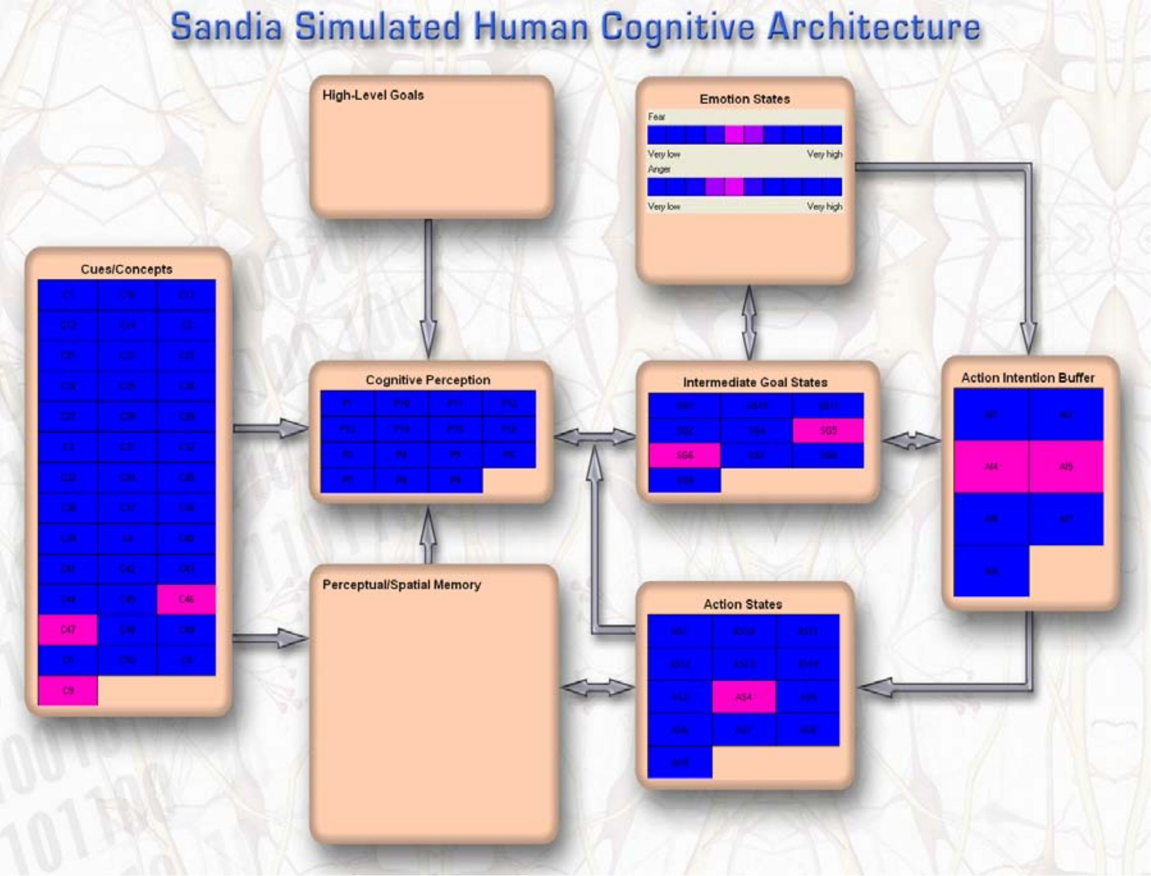


Figure F-3. Sandia Simulated Human Cognitive Architecture

The only part of the software Design Model in Figure F-3 that is not represented in the Figure F-2 is the Perceptual Memory block. This block is represented in the lower left part of the Figure F-1. That part of the Figure F-3 is illustrated by the model fragment shown in Figure F-4 that displays the activation flow of perceptions, spatial memory, and actions for an agent.

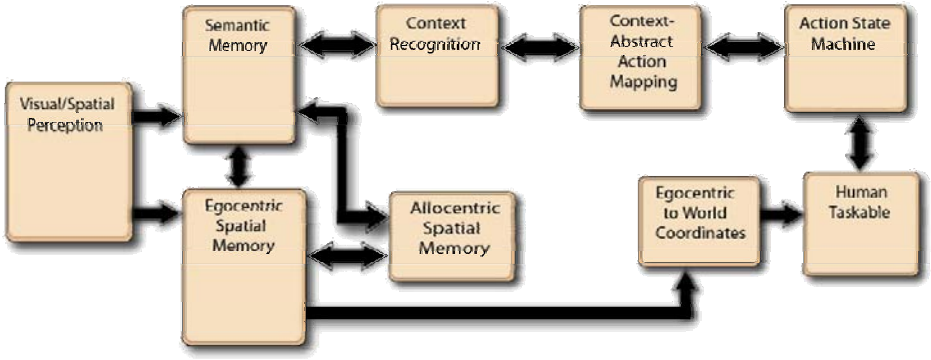

Semantic Memory
Context Recognition
Context-Abstract Action Mapping
Action State Machine
Visual/Spatial Perception
Egocentric Spatial Memory
Allocentric Spatial Memory
Egocentric to World Coordinates
Human Taskable

Figure F-4. Visual_Perceptual Cognition and Action Generation

A more complete representation of the full memory processing with the interactions of these representations with the upper right and lower right models illustrated in Figure F-1 is illustrated in Figure F-5.

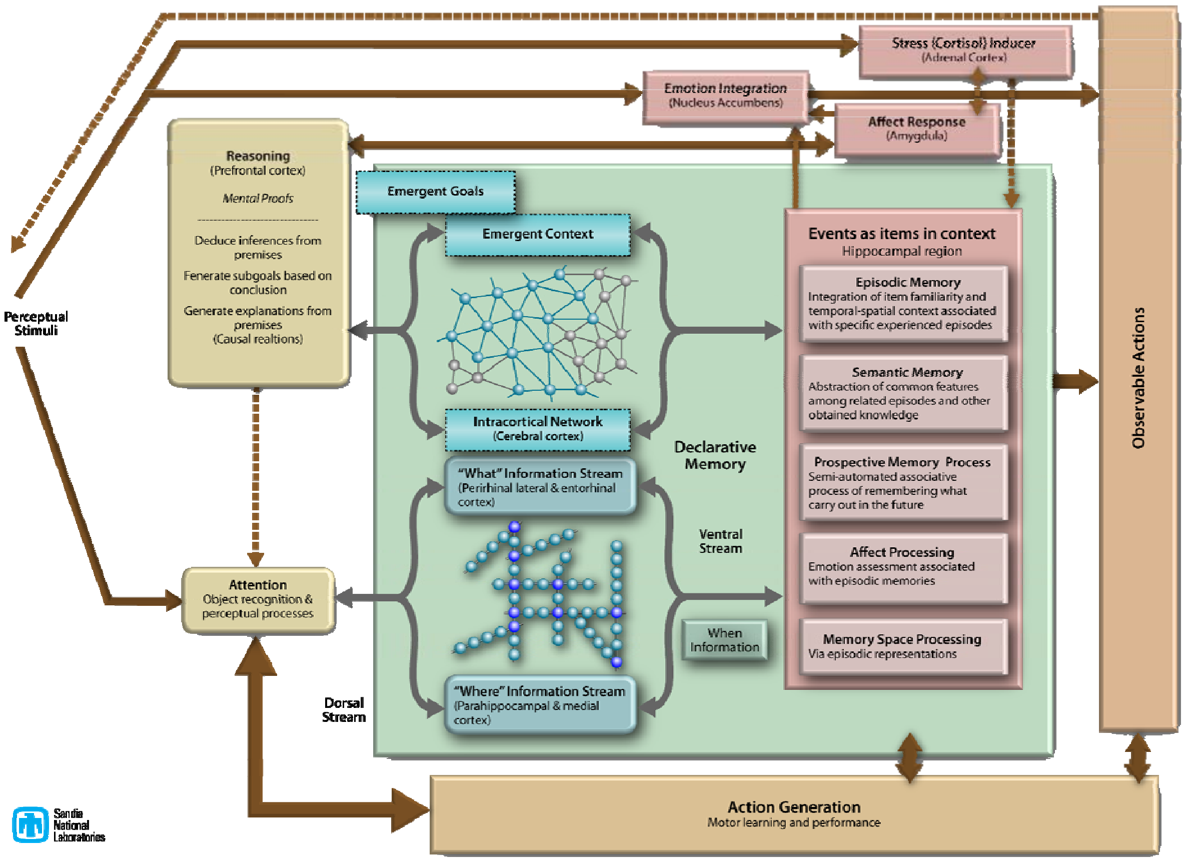

Stress (Cortisol) Inducer
(Adrenal Cortex)
Emotion Integration
(Nucleus Accumbens)
Affect Response
(Amygdula)
Reasoning
(Prefrontal cortex)
Mental Proofs
Deduce inferences from premises
Fenerate subgoals based on conclusion
Generate explanations from premises
(Causal realtions)
Perceptual Stimuli
Emergent Goals
Emergent Context
Intracortical Network
(Cerebral cortex)
Declarative Memory
"What" Information Stream
(Perirhinal lateral & entorhinal cortex)
Ventral Stream
When Information
"Where" Information Stream
(Parahippocampal & medial cortex)
Dorsal Stream
Attention
Object recognition & perceptual processes
Events as items in context
Hippocampal region
Episodic Memory
Integration of item familiarity and temporal-spatial context associated with specific experienced episodes
Semantic Memory
Abstraction of common features among related episodes and other obtained knowledge
Prospective Memory Process
Semi-automated associative process of remembering what carry out in the future
Affect Processing
Emotion assessment associated with episodic memories
Memory Space Processing
Via episodic representations
Observable Actions
Action Generation
Motor learning and performance
Sandia National Laboratories

Figure F-5. Integrated Memory Processing Design Model

This graphic Figure F-5 (in particular the blue layered area is what is emerging from this Human Memory and Reasoning research project and illustrates the continued evolution of the higher fidelity design model from the conceptual design model. As such models are verified and validated and referenced more widely, they may become normative references for their specific area of study. See (Bernard-2008) for more detailed information on this design model.

F.2.4 Case Study Evaluation Model

Some verification and validation test cases and activity have been defined by the project that fall into an Evaluation Model instance for this case study. See Trucano[102] for more detailed information on the V&V Model details applicable to the conceptual Evaluation Model.

1. VERIFICATION TESTING

Verification is the process of determining that a model implementation accurately represents the developer's conceptual description of the model and the solution to the model. The description below is for the Category Verification Test.

[102] T. Trucano, et al. "R&D for Computational Cognitive and Social Models: Foundations for Model Evaluation through Verification and Validation," SAND2008-6453, September 2008.

Framework Note: Although this is called a “verification test”, the conditions (e.g., acceptance criteria for the comparisons/verifications) for defining what has been verified are not stated other than in general terms. Data input scenarios need to be specified to ensure coverage.

• Specification:

The category test is an internally focused analysis of category formation within the individual adaptive resonance theory (ART) modules within the M&R model. Each of the Fuzzy ART1 modules forms self-organized categories corresponding to input data similarities. The analysis of the internal category formation has been selected as a pertinent unit test as it measures the rate of classification within each Fuzzy ART1 module. This rate serves as a meaningful metric in the sense that if a unique category is formed with every presentation of input data, even with repeated presentation of the same input, then no self-organized clustering is occurring, and conversely if a single category is representing all input data, then input data differentiation is not occurring. Additionally, the capability of the model to forget is an internal function of the Fuzzy ART1 modules, and may be observed as a decrease in the number of categories as opposed to an infinite perfect memory which has no bounding capacity and never forgets.

• Referents:

Each ART module will produce a listing of the number of actively committed categories at each time step of the simulation. Similarly, the total number of categories ever formed will also serve as a meaningful metric as it additionally factors in the categories which have been forgotten. Beyond simply measuring how many categories have formed, an additional analysis will track the activation of particular categories as a function of time to determine the distribution of category usage. And finally, a final referent is whether a particular category is reformed after previously having been forgotten.

• Implementation:

This test is implemented as a preprocessor compile option within the software, such that it may selectively be enabled to track categories as desired. Each Fuzzy ART1 module generates a comma separated value (CSV) output file which records pertinent metrics whether it be the number of categories or the distinct weights corresponding to the categories.

The test will be executed repetitively and systematically for each appropriate version of the model functional progression. The resulting output files may then be archived as desired for comparison across version instantiations.

Due to the fact that the referent data is generated as a function of preprocessor enable flags, the test is directly embedded within the code itself which is under version control. Additionally, further processing scripts may be utilized as desired to process the output data offline without affecting system performance.

• Execution:

The testing will be executed on the desired test platform. As a research application as opposed to commercial production software, the code has been developed to be platform independent, but cannot feasibly be tested on every existing variant of operating system and platform. If so desired, the tests may be repeated on alternative platforms with the resulting output analyzed for differences. Discrepancies among the resulting output files would then be cause for concern as an indication of an unintentional difference amongst the implementations resulting from compiler interpretations.

Furthermore, testing may be repeated by varying test parameters such as the input data, the vigilance parameter of an ART module, and whether forgetting is implemented within the model. Clearly the quantity of distinct input patterns as well as their presentation format will influence the formation of categories. Additionally, the vigilance parameter is a property of ART neural networks by which the sensitivity of category differentiation is controlled, and thus varying the vigilance clearly affects the resulting rate of category formation as well.

• Analysis:

When executing the code with the desired test preprocessor flags enabled, CSV files are automatically generated in the root directory the source code is located in. These CSV files are named to correspond with the component labels instantiated in the architecture definition spec files, and depict the referent data. Note: any variation to test parameters require the output files be saved as desired, otherwise repeated execution overwrites the same output files.

• Comparison:

The resulting CSV analysis files are direct embodiments of the referents and may be analyzed directly or if so desired may be further processed for further statistical analysis beyond the code verification of category formation.

• Pass/Fail:

The model as a whole is a heterogeneous combination of components; however the category test only applies to the Fuzzy ART1 modules. The software has been implemented such that rather than requiring distinct code for each Fuzzy ART1 module, the same code may be re-used with appropriate input parameters. The baseline criterion by which the Pass/Fail judgment may be evoked is whether or not a single category has been formed. If no categories are formed, then the Fuzzy ART1 modules are clearly not functioning properly and the system as a whole fails. As the other extreme, if new categories are formed even with repeated presentation of the same input patterns and an appropriate vigilance parameter, then the module also fails. Beyond these extremes, more meaningful judgments are dependent upon the input data as well as Fuzzy ART1 parameters.

• Implication:

The overall implication of the category test is the verification of whether or not memories are being formed within the Fuzzy ART1 modules. The formation and storage of retrievable memories within the individual components comprising the overall architecture is an essential capability to model the corresponding cortical components of human memory. Without establishing the successful functionality of category formation other capabilities which rely upon category formation and storage such as memory recollection or recognition cannot be reasonably implemented or analyzed either.

2. VALIDATION TESTING

Validation is 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. The description below is for the Recall Test.

Framework Note: Although this is called a “validation test”, the conditions (e.g., statistical/probabilistic measures, uncertainty, margins) for defining what has been validated are not stated. Hence, this is really more of a calibration activity – which is the nature of the research effort. Once the prediction parameters (e.g., margins, uncertainty) of the implementation has been established – then to do a statistically valid, independent experiment to see how well the predicted results match with real results would be a more complete validation effort. The “Experimental Validation Plan” might be the focus for the activities of the more complete validation effort. So… there are a couple of “validation” testing activities defined, but there are many more that should be conducted before claiming the resulting model is “validated”.

• Specification:

The recall test is subdivided into top-down recall and spontaneous recall sub-tests. The primary distinction between the two sub-tests is that top-down recall requires manual stimulation of a desired category in the upper Fuzzy ART1 module, whereas the spontaneous recall test is activated by presenting an input cue. The manual probing of the top-down recall test is out of necessity to analyze models without a hippocampus.

Regardless of the stimulation technique required to trigger recall, the significance of testing the recall capability of the model is that without being able to retrieve what has been stored, memory is not useful. Without the capability to verify what is in memory, there is no proof that anything is actually being stored. Consequently, this test pertains to the entire model as whole considering it requires memories must be both stored as well as retrieved.

• Referents:

For both recall tests, first of all the recalled pattern sequence is a referent. Secondly, in relation to the recalled pattern the recall accuracy and ordering are also pertinent referents.

• Implementation:

When recall is initiated the sensory input is no longer received. The category is either manually stimulated or spontaneously activated in the hippocampus component passes down through the lower layers in the model until an output sequence is produced at the input layer.

• Execution:

The desired test cases will be performed on Sandia platforms to generate the referent data, which will further be passed on for comparison study by Neal Cohen. The tests will be repeated in accordance to various input patterns such as the repeated fixed order presentation of basic object-context pairs as well as probabilistic Markov orderings of the possible pairs.

• Analysis:

The results of executing the recall test are sequences of object-context pairs which are associated with the given input cue or the manually probed category.

• Comparison:

The most basic comparison is analyzing whether or not the sequence of object-context pairs recalled is identical to the given input cue in terms of sequential ordering as well as object-context accuracy. Furthermore, the results will also be compared with human subject testing data collected by Neal Cohen. The human subject data will be collected from both healthy and amnesiac patients and so as appropriate the model performance can be validated against both.

• Pass/Fail:

A preliminary judgment of whether the model passes or fails is whether or not the model is able to recall the appropriate sequence in terms of both order and accuracy. Overall however, the pass/fail judgment will be the comparison between the human subject data and the model performance. A perfect computer storage design is not the goal of the model, but rather is intended to model fallible human memory and so perfect recall is not expected under all circumstances.

• Implication:

The overall implication of the recall test is the capability of the model to locate and retrieve stored memories. A validated model is meaningful as a research tool intending to provide the ability to further understand how human memory works. As a predictive measure, a validated model may also be used to analyze to hypothesize the effect of damage to particular cortical regions.

F.2.5 Case Study Risk Model

A sample of some of the risks for each of the framework models are compiled in the Table 5-4 below. A probability index of high, medium or low has been assigned to each threat. This probability index can be used to determine how vital the mitigation plans might be for this project. In general, threats with a high probability should have a mitigation plan in place based on criteria of acceptance.

Framework Note: While a comprehensive summary of risks is not provided at this time due to the early phase of the research, the table below can be used to expose surety gaps and highlight potential gaps that should be addressed in the future. The project’s meeting notes and capture of email correspondence provides an excellent resource for detailing a risk model.

Table F-1. Case Study Risk Model Summary

Vulnerability	Threat	Likelihood	Potential Consequence	Impact	Current Practices	Gap?	Mitigation Plan
New computational models are validated per current psychological literature	Future psychological research results do not correspond with computational models	Medium	**Computational models will not prove valid in the future**	Low		Yes	If there are major competing theories for a particular cognitive function, we will test and potentially model each theoretical version.
Experimental controls for recall data measurements might be effected by external stimuli	Human subjects perform gaze aversion	High	**Outside intervention might effect data results for recall**	High		**Yes**	
Measurements are conducted in a controlled environment	Model behaves autonomously	Medium	**Uncertainty for measuring a model that behaves on its own**	Medium		**Yes**	
Human subjects' eye tracking data will be collected	Data collected from physiological measurements are not safeguarded	High	**Human subjects' privacy might be compromised**	High	HSB review and approval	**No**	

F.2.6 Case Study Maturity Model

The Cognitive System Maturity Matrix in Table F-5 provides an at-a-glance view of the development level of the fidelity of the Modeling Aspects of Human Memory and Reasoning project. Overall, this project is clearly in the research/experimental stage where understanding of the concepts is the focus. There is limited attention at this time to the consideration of an array of environments, surety, risks, and any ethical, legal, and social implications as summarized below.

Table F-2. Cognitive System Maturity Matrix

Fidelity Level / Attribute	Level 0 Low Consequence, Minimal Impact Scoping Studies & Research Models for Understanding	Level 1 Moderate Consequence, Some Impact Preliminary Product Experimental Use	Level 2 High Consequence	Level 3 High Consequence, Qualified/Certified
Psychological Representation	▪ Theories are in developmental stage	▪ Major elements are represented by behavioral models in psychological literature (ref Eichenbaum, McClellan)		
Physiological Representation	▪ Theories are in developmental stage	▪ Physiological measurements using calibrated test equipment. ▪ Models are empirical with no basis for extrapolation to other applications.		
Environmental Representation	▪ Limited normal scenarios defined. ▪ No abnormal scenarios identified. ▪ No hostile scenarios identified.			
System Surety Engineering	▪ Limited to laboratory environment. ▪ Uncertainty of data measurements are expected to have a large effect on results ▪ Theory is in development	▪ IRB approval obtained, privacy issues addressed ▪ Experimental reliability of results will be analyzed for critical parameters ▪ V&V plan developed		
Ethical, Legal and Societal Implications	▪ Limited discussions	▪ IRB approval obtained, privacy issues addressed		
System Risk Mitigation	▪ Significant research gaps identified	▪ Large and unknown uncertainties in experimental evidence and tests expected		

APPENDIX G - PROJECT LEAD BIOGRAPHIES

G.1 David E Peercy, PhD, SNL

DAVID PEERCY is a Distinguished Member of the Technical Staff at Sandia National Laboratories in the Weapon Systems and Software Quality department. Dave received his PhD in Mathematics from New Mexico State University. His current work focuses on quality engineering of weapon systems and software, with a particular emphasis on the associated surety technologies. Dave has developed international standards and guides in the areas of software reliability, software supportability, and software safety. He is the principal investigator on the research presented in this report as well as a research project to develop a methodology for the quantification and measurement of the quality of nuclear weapon and weapon-related systems design.

G.2 Wendy Shaneyfelt, SNL

WENDY SHANEYFELT is Computer Science Researcher contracting to Sandia National Laboratories in the Cognitive System Research and Applications department. She received her BS in Computer Science from the University of Nebraska. Her current work involves applying cognitive system technologies towards analytical tasks to significantly increase effectiveness and efficiency. Wendy developed ethical principles and guidelines for the development of cognitive systems and is actively involved in exploring the societal, legal, ethical, and political implications surrounding cognitive system technologies. As a key researcher on the project presented in this report she applies the surety framework developed herein to current government projects.

G.3 Eva Caldera, JD, UNM

EVA CALDERA is Research Professor of Law, University of New Mexico School of Law, and Associate Director of the Institute for Ethics, UNM Health Sciences Center. She received both her undergraduate degree in Philosophy and her law degree from Harvard University. Her current work involves teaching and research in bioethics and ethics of emerging technologies, and includes courses taught in law and bioethics, public health law and ethics, and ethics of nanotechnology. She has worked on several projects with Sandia National Laboratories directed at the analysis of legal and ethical issues associated with the development of cognitive systems. Her other research includes an NIH-supported project to develop a tele-health model to provide training in clinical ethics consultation throughout the state of New Mexico. She serves on the Human Research Review Committee for the UNM Health Sciences Center and on the Biomedical Ethics Committee for UNM Hospital.

G.4 Thomas Caudell, PhD, UNM

THOMAS CAUDELL received a bachelor’s degree in Physics and Mathematics from California State University at Pomona in 1973, and a master’s degree and a PhD in Physics from the University of Arizona, Tucson, in 1978 and 1980, respectively. From 1984 through 1989, he was a Senior Staff Physicist at the Hughes Artificial Intelligence Center, Hughes Research Laboratories, in Malibu. From 1989 through 1993, he was a Senior Principal Scientist in Research and Technology at Boeing Computing Services in The Boeing Company in Seattle. Dr. Caudell came to UNM in 1994 and is now a tenured Full Professor of Electrical and Computer Engineering and Computer Science at the University of New

Mexico in Albuquerque, and Director of the University of New Mexico Center for High Performance Computing.

Professor Caudell’s general areas of research are in Computational Cognitive Neural Sciences and Advanced Human-Computer Interfaces. Specifically his research program is addressing a) cognitive and neuro-biologically motivated mathematical theories and simulations of neural systems, both biological and artificial, and b) virtual reality/game environments to enhance human comprehension of complex phenomena such as neural systems. His research program is highly interdisciplinary, involving collaborations with neuroscientists, psychologists, physicians, mathematicians, computer scientists, artists and musicians.

DISTRIBUTION

32 Sandia National Laboratories Internal Distribution

#	**Mail Stop**	**Name, Organization**
1	MS 0122	Gerold Yonas, 00700
1	MS 0370	Timothy Trucano, 01411
1	MS 0370	Laura McNamara, 01433
1	MS 0415	Ron Pedersen, 00241
1	MS 0428	Rick Fellerhoff, 12300
1	MS 0428	Todd Jones, 12330
1	MS 0428	Michael Daily, 12340
1	MS 0487	Brian Geery, 02113
1	MS 0637	Ricardo Sarfaty, 12336
1	MS 0638	Ron Farmer, 12341
5	MS 0638	David Peercy, 12341
1	MS 0757	John Russell, 06414
1	MS 1011	Jonathan McClain, 06343
5	MS 1011	Wendy Shaneyfelt, 06343
1	MS 1011	Ann Speed, 06343
1	MS 1188	Gerard Sleefe, 06340
1	MS 1188	Michael Bernard, 06341
1	MS 1188	Kevin Dixon, 06341
1	MS 1188	Chris Forsythe, 06341
1	MS 1188	John Wagner, 06341
2	MS 9018	Central Technical Files, 8944
2	MS 0899	Technical Library, 04536

4 University of New Mexico Distribution

ATTN: Eva Caldera (2)
ATTN: Tom Caudell (2)
Center for High Technology Materials
MSC04 2710
1313 Goddard SE
Albuquerque, New Mexico 87106-4343

Sandia National Laboratories