SANDIA REPORT
SAND2006-6895
Unlimited Release
Printed November 2006

Investigating Surety Methodologies for Cognitive Systems

David E. Peercy and Wendy L. Shaneyfelt, Eva O. Caldera,
Tom P. Caudell, and Kirsty Mills

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.

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

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

SAND2006-6895
Unclassified Unlimited Release
Printed November 2006

Investigating Surety Methodologies for Cognitive Systems

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

Eva Caldera
Associate Director, Institute for Ethics
Tom Caudell
Director, Center for High Performance Computing Visualization Laboratory
Kirsty Mills
Associate Director, Center for High Technology Materials
University of New Mexico
1313 Goddard SE,
Albuquerque, NM 87106

Abstract

Advances in cognitive science provide a foundation for new tools that promise to advance human capabilities with significant positive impacts. As with any new technology breakthrough, associated technical and non-technical risks are involved. Sandia has mitigated both technical and non-technical risks by applying advanced surety methodologies in such areas as nuclear weapons, nuclear reactor safety, nuclear materials transport, and energy systems. In order to apply surety to the development of cognitive systems, we must understand the concepts and principles that characterize the certainty of a system's operation as well as the risk areas of cognitive sciences.

This SAND report documents a preliminary spectrum of risks involved with cognitive sciences, and identifies some surety methodologies that can be applied to potentially mitigate such risks. Some potential areas for further study are recommended. In particular, a recommendation is made to develop a cognitive systems epistemology framework for more detailed study of these risk areas and applications of surety methods and techniques.

ACKNOWLEDGEMENTS

The authors would like to thank the many anonymous individuals who generously gave their time to participate in our study groups. Your valuable insights have contributed to furthering the responsible design and development of cognitive science technologies.

In addition, special gratitude is expressed to the Center for High Technology Materials (CHTM) and the Mental Illness and Neurosciences Discovery (MIND) Institute at the University of New Mexico for providing support personnel, facilities for focus group interviews, refreshments, recording equipment, and many other amenities that greatly facilitated the conduct of this short research effort.

Table of Contents

ACKNOWLEDGEMENTS **4**

EXECUTIVE SUMMARY **7**

1. INTRODUCTION **13**

1.1. PURPOSE 13
1.2. SCOPE 13
1.3. PARTICIPANTS AND AUDIENCE 14

2. TECHNICAL APPROACH **15**

2.1. BACKGROUND 15
2.2. COGNITIVE SCIENCE AND TECHNOLOGY 15
2.3. SURETY PRINCIPLES AND METHODS 16
2.3.1. Safety *16*
2.3.2. Reliability *16*
2.3.3. Security/Use Control *17*
2.3.4. Human Factors *17*
2.3.5. Quality *18*
2.3.6. Surveillance – System Sustainment *20*
2.4. RESEARCH APPROACH 20
2.4.1. Questionnaire Survey and Reference Basis *20*
2.4.2. Focus Groups *27*

3. SUMMARY OF RESEARCH RESULTS **29**

3.1. FOCUS GROUP DEMOGRAPHICS 29
3.2. COGNITIVE TECHNOLOGY IMPACT ON SOCIETAL AND ETHICAL ISSUES 32
3.3. APPLICATION AREA RISKS 33
3.4. SCIENTIST RESPONSIBILITIES 35
3.5. FOCUS GROUP DISCUSSION HIGHLIGHTS 39
3.5.1. Reliability *39*
3.5.2. Privacy *40*
3.5.3. Liability *40*
3.5.4. Legal/Ownership/Intellectual Property *41*
3.5.5. National Security *42*
3.5.6. Hype and Backlash *42*
3.5.7. Dependency *43*
3.5.8. Diversity *43*
3.5.9. Equity *44*
3.5.10. Human Enhancement *44*
3.5.11. Moral/Religious/Spiritual *46*
3.5.12. Other Interesting Comments *46*
3.6. SPECTRUM OF RISKS 50
3.7. SURETY METHODS AND APPLICABLE AREAS 52

4. CONCLUSIONS AND RECOMMENDATIONS **55**

4.1. KEY CONCLUSIONS 55
4.2. KEY RECOMMENDATIONS 56

APPENDIX A - REFERENCES **57**

APPENDIX B - GLOSSARY **58**

B.1 ACRONYMS 58
B.2 DEFINITIONS 58

APPENDIX C - PROJECT LEAD BIOGRAPHIES **62**

C.1 David E Peercy, PhD, SNL 62
C.2 Wendy Shaneyfelt, SNL 62
C.3 Eva Caldera, JD, UNM 62
C.4 Tom Caudell, PhD, UNM 62
C.5 Kirsty Mills, PhD, UNM 63

APPENDIX D - RESEARCH DATA 64

D.1 Focus Group Interview Materials 64
D.1.1 Participant Consent Form and Survey Questionnaire 64
D.1.2 Focus Group Interview Session Procedure 64
D.1.3 Pre-Meeting Participant Reading Material 65
D.2 Workshop and Conference Discussion Data 65
D.2.1 Poster Presentation: Where Do You Draw the Line? 65
D.2.2 Cognitive Dominance Workshop 65
D.3 Focus Group Questionnaire Survey Data 66
D.4 Focus Group Discussion Recording Data 66
D.4.1 Summary Transcript Across All Sessions 66
D.4.2 Non-Technical Focus Group Session Transcript 66
D.4.3 Public Focus Group Session Transcript 66
D.4.4 Technical Focus Group Session Transcript 66
D.4.5 Surety Focus Group Session Transcript 66

DISTRIBUTION 67

List of Figures

Figure 2.3-1. Quality Methodology Framework 19
Figure 3.1-1. Focus Group Age Profile 29
Figure 3.1-2. Focus Group Gender Profile 30
Figure 3.1-3. Focus Group Daily Activity Profile 30
Figure 3.1-4. Focus Group Cognitive Technology Expertise Profile 31
Figure 3.2-1. Cognitive Technology Likelihood to Cause Societal and Ethical Issues (By Focus Group) 32
Figure 3.2-2. Cognitive Technology Likelihood to Cause Societal and Ethical Issues (All Focus Groups) 32
Figure 3.3-1. Summary of Perceived Application Area Ethical Risks (All Groups) 33
Figure 3.4-1. Cognitive Scientist Responsibility for Societal Results of Research (All Groups) 35
Figure 3.4-2. Ease of Cognitive Scientist Recognizing Ethical Issues (All Groups) 36
Figure 3.4-3. Cognitive Scientist Responsibility to Public: Explain Research (All Groups) 36
Figure 3.4-4. Cognitive Scientist Responsibility to Public: Discuss Research (All Groups) 37
Figure 3.4-5. Cognitive Scientist Responsibility to Public: Consider Public Opinion (All Groups) 37

List of Tables

Table E-1. Focus Group Demographics 7
Table E-2. Cognitive Technology Societal and Ethical Issue 7
Table E-3. Cognitive Technology Application Area Ethical Issues 8
Table E-4. Cognitive Scientist Responsibility and Ethical Issues 8
Table E-1. Cognitive Science Technology Risk Areas & Applicable Surety Methods 8
Table 2.4-1. Summary of Questionnaire Survey 20
Table 2.4-2. Summary of Risks, Fears, and Potential Mitigation Strategies Informally Provided by Workshop Participants. 21
Table 2.4-3. Highlights from Cognitive Dominance Ethics Workshop 22
Table 2.4-4. Focus Group Interview Procedure 28
Table 3.6-1. Preliminary Perceived Risk Spectrum 51
Table 3.7-1. Perceived Risk Spectrum With Applicable Surety Methods 52

EXECUTIVE SUMMARY

This report describes the results of the Investigating Surety Methodologies for Cognitive Systems Lab Directed Research and Development (LDRD) project, #105306. The purpose of this LDRD was to identify a spectrum of risks associated with cognitive systems and investigate the application of surety methods to potentially mitigate some of these risks.

Technical Approach

Sandia National Laboratories teamed with the University of New Mexico to attend workshops, conduct Focus Group surveys and interviews, distribute a questionnaire survey, and study existing publications to obtain a perspective on potential risks of cognitive systems. The initial scope of Cognitive Systems was targeted at the following definition:

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

However, the project discussions covered a broader definition of cognitive systems ranging from augmenting human cognition to simulating human cognitive tasks. The systems considered were integral to the human's cognitive tasks, either externally or internally to the human body. Surety specialists from Sandia National Laboratories provided suggestions as to what surety methods might be applied to mitigate the identified risks.

Results

Demographics of the Focus Groups are summarized in Table E-1.

Table E-1. Focus Group Demographics

Age Profile: biased toward "middle age"
Gender Profile: biased toward "Male"
Daily Work Profile: equally distributed among Technology Management, Technology R&D, and non-Technical activities
Cognitive Science Experience: equally distributed "Not Expert" to "Very Expert"

The general response to whether cognitive technology would cause societal and ethical issues is summarized in Table E-2.

Table E-2. Cognitive Technology Societal and Ethical Issue

Cognitive Technology Ethical Issue: there is a strong perception by each Focus Group and across all Focus Groups that it is likely cognitive systems would cause societal and ethical issues.

Nine application areas were identified in the questionnaire survey. Respondents were asked whether cognitive systems posed ethical issues for these nine application areas. The results are summarized in Table E-3.

Table E-3. Cognitive Technology Application Area Ethical Issues

Application Area Ethical Issues: The application areas of Medicine, Privacy, Human Enhancement, Law & Policy, and Military are of highest concern, although few participants felt any area was not likely to pose ethical issues. Ethical issues drive risks in many of the risk dimensions and specifically the perceived risk spectrum.

One area of potential concern that the questionnaire survey addressed was the responsibility of the cognitive scientists for the eventual societal results (both intended and unintended) of their research, how easy it would be for the cognitive scientists to recognize potential ethical issues associated with their projects, and the responsibility of cognitive scientists to communicate information concerning their projects with the public. The overall results are summarized in Table E-4.

Table E-4. Cognitive Scientist Responsibility and Ethical Issues

Overall Societal Responsibility: strong perception that the cognitive scientist is responsible for societal results (intended and unintended). Minority felt not responsible.

Ease of Recognizing Ethical Issues: there is a fairly equal distribution from "Not Easy" to "Very Easy" concerning the question whether it would be easy for a cognitive scientist to recognize ethical issues in their research. Disagreement may be due to perception of issues being application-dependent, and also to varying degrees of sensitivity among individuals towards potential issues.

Responsibility to Public: strong agreement that the cognitive scientists are responsible to explain and discuss their research with the public. Less agreement on whether the cognitive scientists should be responsible for considering public opinion when deciding the future direction of their research.

The cognitive science technology risk spectrum and applicable surety methods identified in Table E-1 were identified through a series of workshops, Focus Group interviews, and related publications. The data was distilled down to 11 discrete risks as listed in the first column. The rationale, descriptions, and concerns surrounding these risks were consolidated and reported in the second column. The Surety Focus Group studied these identified risks, discussed the concerns, and provided potential surety methods their experience suggested would have potential impacts (third column). In particular, privacy, legal/ownership/IP, and reliability of the systems seemed to be of highest priority.

Table E-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 neuroscience, which currently lacks a broad	Safety Principles Reliability: FMEA/FTA/PM/HF Methods/Sensitivity Analysis Risk Analysis: QMU Quality Methodology Ongoing monitoring efforts to detect adverse consequences early.

	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	Redundancy, system backups, and high reliability in systems will be crucial to

	technology, and may 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 wellbeing. 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.

Conclusions

From the research results the following seven conclusions have been formulated:

Conclusion 1: There was general consensus among the project Focus Groups that there are many areas of risk in the development and use of cognitive systems. In particular, privacy, legal/ownership/IP, and reliability of the systems seemed to be of highest priority. However, the risks and their priorities were clearly dependent on the intended application.

Conclusion 2: These high priority risk areas of privacy and reliability are particularly susceptible to surety analysis, through techniques such as encryption technology and reliability methodologies.

Conclusion 3: There is a strong perception by each Focus Group and across all Focus Groups that it is likely cognitive systems would cause sociological and ethical issues. The application areas of Medicine, Privacy, Human Enhancement, Law & Policy, and Military are of highest concern. Ethical issues drive risks in many of the risk dimensions and specifically the perceived risk spectrum.

Conclusion 4: There is a distribution of opinion as to how easily cognitive scientists would be able to recognize ethical issues in their research. This may be because the responses were given with various applications in mind, and/or to varying sensitivity to potential issues among respondents. There is strong agreement that cognitive scientists have a responsibility to explain and discuss their research with the public.

Conclusion 5: There is general agreement that cognitive scientists have a responsibility to consider the societal outcomes of their research. There is less agreement on whether the cognitive scientists should be responsible for considering public opinion when deciding the future direction of their research. Cognitive scientists participating in the project Focus Groups gave clear evidence of their concerns over a range of these issues. They also reflected a sense that other societal institutions, including the legislature and the judiciary, will play a major role in this process.

Conclusion 6: Surety methods provide assurance that a system is safe, reliable, secure, built with human factors considerations, and with an appropriate level of quality for the intended application use. Clearly, these surety methods are applicable to specific aspects of cognitive system design, and if appropriately applied would reduce the technical risks associated with such systems. Other risk dimensions, such as the appropriate distribution of the benefits and burdens of technology, for example, are less susceptible to resolution through surety methodology, and fall into the area of public policy and democratic decision-making. The development of reliable and safe technologies, however, will greatly facilitate this public policy debate.

Conclusion 7: The Focus Groups were designed as a pilot study to indicate the qualitative dimensions of the risks associated with cognitive systems. The groups broadly agreed on many of the queried issues, irrespective of group make-up. Future efforts will be designed with attention to diversity across age, gender, and work background.

Recommendations

Furthering the conceptual research information presented in this report and implementing it is imperative to the responsible development of cognitive systems. In order to maintain our leadership position as researchers who responsibly focus upfront on identifying, assessing, and mitigating potential risks involved with the furthering of cognitive science we must remain proactive. There are two recommendations from this short LDRD research study.

Recommendation 1: It is recommended that the initial research results of this LDRD study be published in a recognized journal to acquaint the research committee with this effort.

Recommendation 2: It is recommended that a follow-on LDRD study be initiated to investigate and develop a cognitive systems epistemology framework, integrating within this framework the risk areas/issues, applicable risk dimensions, and surety methods identified in this preliminary study.

This page blank except for this statement.

1. INTRODUCTION

1.1. Purpose

The interest in cognitive science and technology is growing among a global scientific community, as well as United States policy makers. While the advancements of this science appear to promise significant enhancements to humans, the specific risks associated with it are largely unknown. Sandia is uniquely poised to apply risk management strategies to the development and deployment of cognitive systems. By leveraging expertise in applying surety methodologies to many other technical disciplines such as nuclear weapons, nuclear reactors, nuclear material transports, and energy infrastructures, Sandia has the opportunity to establish global leadership in surety science for cognitive systems. Successful results will enable Sandia to contribute to the emerging national initiative in Neurotechnology being advanced by the Potomac Institute, Defense Advanced Research Projects Agency (DARPA), National Science Foundation (NSF) and Office of Naval Research (ONR).

The intent of this short two-month Laboratory Directed Research and Development (LDRD) project on *Investigating Surety Methodologies for Cognitive Systems* was to identify a spectrum of risks associated with cognitive science technologies and examine appropriate surety methods and relevant approaches that might be effective in mitigating these risks. The initial project scope was targeted at a somewhat limited category of cognitive systems, defined as:

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

However, the project discussions covered a broader definition of cognitive systems ranging from augmenting human cognition to simulating human cognitive tasks. The systems considered were integral to the human's cognitive tasks, either externally or internally to the human body.

The results reported do provide a preliminary characterization of some risks involved with cognitive systems and participants' perceptions of whether ethical responsibilities are likely for such scientific investigation. In addition, some surety methods have been identified that might be appropriate to reduce perceived and actual risks associated with the development and application-use of cognitive systems. Preliminary conclusions and recommendations are provided, in particular, a recommendation is made to develop a cognitive systems epistemology framework for more detailed study of these risk areas and applications of surety methods and techniques.

1.2. Scope

Researchers from Sandia and the University of New Mexico (UNM) collaborated on this project to characterize a range of actual, potential, and perceived technical and non-technical risks associated with the development and use of cognitive systems. Areas of potential applications for cognitive systems technologies include:

- Detection, recognition, analysis, and forecasting of human behavior and performance

- Machine representation and application of human knowledge and experience (synthetic subject matter expert)
- System adaptation to the knowledge, skill, situation awareness, or intentions of individual operators or teams of operators
- Preservation and transfer of knowledge and experience
- Aides to human attention, memory, situation awareness, decision-making, and other cognitive functions
- Technologies in which human-machine interaction are vital to the performance, safety, and security of systems
- Training for jobs or tasks in which human interaction under unpredictable and stressful conditions is essential to success

The scope of this short study effort was to characterize a spectrum of potential risks due to the applications of cognitive technologies/systems and determine which if any existing surety methods might be applicable for risk aversion/reduction. This characterization was accomplished in a two-fold approach. Focus groups were established and solicited for their perceptions of the potential risks and surety specialists were queried as to potential surety methods that might be applicable.

Focus groups of individuals were identified from which to solicit ideas and feedback. Focus groups were categorized as normative reference (baseline), technical expert, non-technical expert, general public, and surety technology. A simple questionnaire survey was developed and prototyped for use by the baseline focus group, each individual completing the questionnaire survey anonymously. Each subsequent focus group participated in a 90-minutes interview session: each individual completed a consent form to be recorded during the interview, anonymously completed the questionnaire survey, and participated in an extensive discussion session with the whole group. The discussion sessions were recorded with participant names deleted from the compiled session reports. A comparative analysis of the results from the discussions and survey results across questionnaire survey responses and group discussions provided potential indicators of high-risk areas. Experts in the surety technology group were further tasked to explore how surety technologies could form the basis of potential science and engineering tools that could be appropriately applied to cognitive systems to reduce risks.

1.3. Participants and Audience

The audience for this report includes policy makers, cognitive scientists, surety engineers, and computational modeling experts who have an interest in understanding the spectrum of risks involved in the development and ethical use of cognitive systems, as well as potential safeguards to consider.

The participants in the study whose results are presented in this report include research personnel from Sandia National Laboratories and UNM who have expertise in cognitive science and surety technologies. In addition, there were participants selected from other areas within UNM and within the city of Albuquerque who provided their perceptions on the risks of cognitive systems across a broad spectrum of applications via discussions and questionnaires.

2. TECHNICAL APPROACH

2.1. 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 – reliability, safety, security, use control, surveillance, human factors, and quality – have been in place for multiple decades. Recognizing both known and unknown risks were involved while pursuing the cognitive systems research and developing technologies, an obvious prudent step mandated understanding the risks and then determining deliberate methods to mitigate them. Working in partnership with the 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.

2.2. Cognitive Science and Technology

The purpose of Sandia's Cognitive Science and Technology (CS&T) Program, established in 2006, is to create 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, Weapons of Mass Destruction (WMD) proliferation, and social unrest due to disruptive forces from changing societies, economies, and climate. The three CS&T objectives are: 1) a basic science understanding of human behavior, 2) improved human performance, and 3) advanced human-machine systems at all scales ranging from nano to social.

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 machines that have 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 technologies augmenting human decision-making, information overload, knowledge preservation, and training.

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. However, concerns have been raised pertaining to such issues as the individual's privacy, legal ramifications for a model's use, and the verification and validation of the model.

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 the overarching quality practices can potentially reduce the risk (perceived or actual) so that the cognitive system might be considered acceptable for use (in some cases even development) and its use validated for specific applications.

2.3. Surety Principles and Methods

The surety areas of interest include: safety, reliability, security, human factors, and quality. Surveillance is another area of interest, primarily in terms of sustainment of a system. These areas have been significantly studied and applied within the SNL weapon/weapon-related applications as well as for other technologies. This section briefly describes some of the principles in these areas and typical methods that might be used to assure models and their computational implementation. Application of such principles and methods may result in the reduction of risk associated with cognitive systems. A general relationship between application of surety principles and methods and potential reduction in the perceived cognitive system risk spectrum is described in Section 3.6.

2.3.1. Safety

Sandia has developed a strong infrastructure and process definition [DG10100-2003] that ensures systems are safe. Cognitive systems must also exhibit a strong 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 dependent on the behavior of such systems. For safety purposes, cognitive systems should attempt to be developed using the following four principles of isolation, independence, inoperability, and incompatibility.

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 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.

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.

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) [TRUCANO-2006] 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
- 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. Quality

Quality is essentially the management of vulnerabilities to a targeted risk. When systems have limited vulnerabilities that can be exploited by threats, the system will have a high level of 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/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. One current project that holds much promise for ensuring weapon/weapon-related product quality is the development of a Quality Methodology for Quantifying and Measuring the Quality of Nuclear Weapons Design. A generic architecture has been developed that would also apply to any system – in particular a cognitive system. Four models are part of the architecture: Specification Model, Design Model, Evaluation Model, and Quality Model as illustrated in the Figure 2.3-1. Key terminology for this model includes:

- Use Case Scenario: Sequence of activities through which the weapon (system, subsystem, component) part is intended to satisfy its specifications in accordance with how it has been designed
- 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
- Specification Model: Generic requirements (behavioral, structural, environmental) that address the class of products/processes within the scope of the quality framework

  Instance: requirements of a specified product/process
- Design Model: Generic architecture (physical/functional) that describes the class of products within the scope of the quality framework

  Instance: design and processes used for a specified product

- Evaluation Model: Generic processes and methods that might be used to obtain measures of how well the generic requirements of the specification model are met by the generic architecture of the design model

  Instance: specific processes and methods used to obtain measures of how well the requirements of a specified product/process are met

- Quality Model: Generic analysis processes and methods that might be used in a time/phase-dependent approach to determine the quality implications of the measures obtained by the Evaluation Model

  Instance: time-dependent identification of gap (potential vulnerability) indicators from the Evaluation Model instance, risk-based analysis (potential threat, impact of threat/vulnerability occurrence, and likelihood of occurrence) of the gap indicators, and representation of the gap indicators in a risk-based prioritization across the life cycle

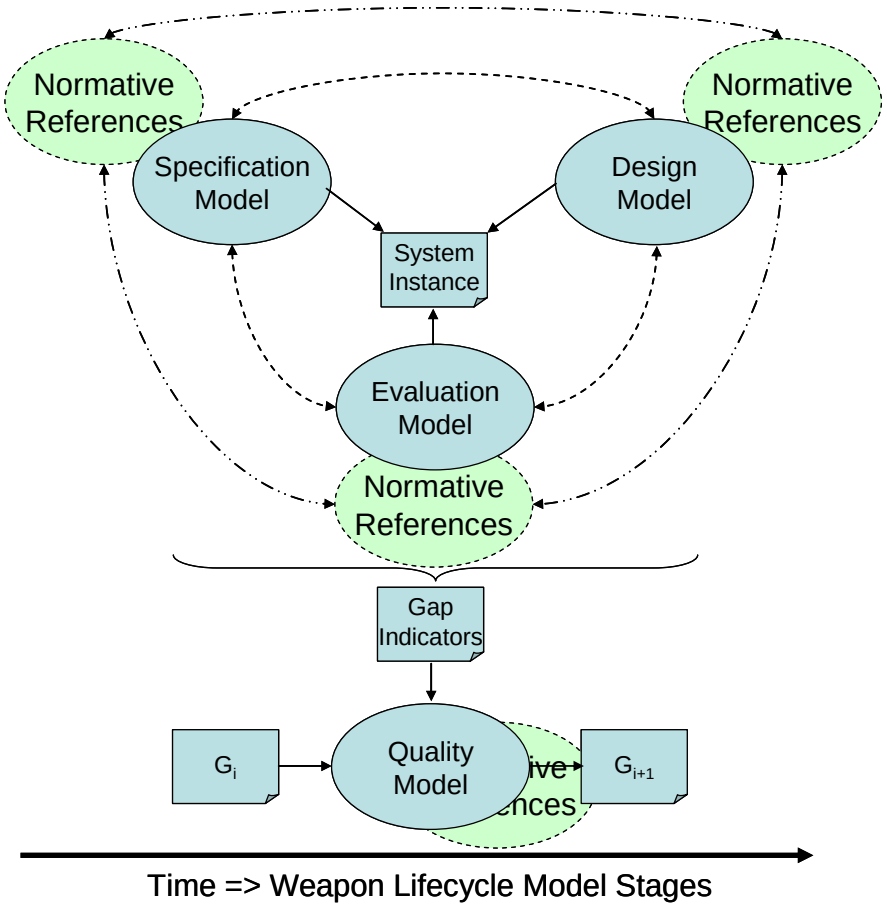

Figure 2.3-1. Quality Methodology Framework

2.3.6. 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.

2.4. Research Approach

The research approach selected had to be appropriate for the short duration (approximately two months) of the project. Preliminary work in the cognitive science area was to be used as a reference, planned workshop and conference presentations were to provide immediate input to the study, and stratified focus groups were to be organized and interviews conducted. A general purpose questionnaire survey was developed and prototyped for use by the focus groups. An interview procedure was developed so as to minimize procedural differences among the focus group interviews. The intent was to solicit opinions from the focus groups as to their perceptions of cognitive systems' risks. Although not intended to provide a strict sample-size, statistically valid set of focus groups, the intent was to determine if somewhat different groups of people had similar or different perceptions of the risk areas.

2.4.1. Questionnaire Survey and Reference Basis

2.4.1.1. Questionnaire Survey

The questionnaire survey consisted of the ten questions summarized in Table 2.4-1, with appropriate Likert answer scales (mostly 1 to 5 from least to most unless otherwise indicated). Comments were also solicited throughout the questionnaire and are included in the full questionnaire survey artifact and survey results referenced in Appendix D.

Table 2.4-1. Summary of Questionnaire Survey

1. How **expert** would you consider yourself to be in cognitive technology?
2. Are your daily activities: a. In technology management? B. In technology R&D? c. Not in the area of technology?
3. How **probable** do you think it is that cognitive technology will pose societal and ethical issues?
4. How **likely** do you think it is that cognitive technology will pose ethical issues in the following areas? o Medicine o Privacy o Human enhancement o Health and Safety o Law and Policy o Economics o Education o Military o International o Other
5. How **responsible** do you think scientists should be for the eventual societal results (both intended and unintended) of their research?
6. Suppose that you are a scientist working on a project in cognitive technology. How **easy** do you think you would find it to recognize ethical issues associated with your project?
7. How **responsible** do you think a scientist is to do the following: o **Explain** his or her work **to** the general public? o **Discuss** his or her work **with** the general public?

o **Consider** public opinion when deciding the future direction of his or her work?
8. How old are you? o < 18 o 18-22 o 23-26 o 27-30 o 31-40 o 41-50 o 51-60 o > 60
9. What is your gender? o Male o Female
10. Are there any comments you would like to add?

2.4.1.2. Reference Basis

Three conference/workshop sessions contributed to the reference basis for this LDRD. Also, a public response to a cognitive systems article authored by a Sandian was captured from an Internet website

2.4.1.2.1. Cognitive Workshop Poster Session Summary

The *2006 Cognitive Systems Workshop: Bridging Cellular to Social* was held in Santa Fe, NM on June 27-29. Eva Caldera and Wendy Shaneyfelt presented the poster "Where Do You Draw the Line?" The objectives of this poster presentation were 1) to stimulate discussion on the spectrum of risks associated with cognitive systems technologies, and 2) to elicit feedback from members of the cognitive systems community describing both real and perceived risks. To achieve these objectives, case scenarios were presented in the poster presentation with white space provided for informal comments. A questionnaire survey was developed and distributed to provide a more formal method of feedback. This survey sample served as our Baseline focus group. This same questionnaire survey was later distributed to the other focus groups.

A summary of the comments from the workshop participants concerning risks, fears, and potential mitigation strategies is provided in the Table 2.4-2. A more detailed description of the Case Scenarios studied and the comments captured can be found in Appendix D.

Table 2.4-2. Summary of Risks, Fears, and Potential Mitigation Strategies Informally Provided by Workshop Participants.

Risk Area	Fear	Potential Mitigation Strategies
Individual rights	What if we don't understand what we're signing up for?	Informed consent is necessary to explain risks.
	What if we don't want to be cognitively enhanced? Will we be forced to? Will it be expected of us from our employer? Are we, the public, being manipulated?	Right to refuse.
Humanity	Can this be used to change behavior?	

	Can this be used forcefully to change behavior (as in controlling/monitoring prisoners?) Will we get to choose the degree of enhancement desired? Will cognitive enhancements give us "rite of passage"?	
Not pursuing technology	What if our enemies are progressing with these technologies and we're not? Can we understand the impact if we're not pursuing the science ourselves?	Pursue science.
Health	Degradation of device in brain. Infection. Can it be removed safely and are there any side effects to removing it (e.g., physical, psychological, emotional, etc?)	Enhancements should be reversible.
	Can it be tampered with by self, enemies? Can it be destroyed while embedded in the brain (perhaps due to an EMP?)	Devices should be failsafe and anti-tamper proof.
	Can its operation be controlled? Is it tunable to turn certain features on/off, and if so, by whom?	On/off switch, tuning mechanisms.
	What are the effects of de-enhancement?	Focus on psychological aspects of human enhancement.
	What upgrades/maintenance will be required?	
Child Protection	Are there age restrictions? Are there considerations made to the maturity of the brain and capability to discern/judge?	

2.4.1.2.2. Cognitive Dominance Ethics Subgroup Summary

The 2nd annual Cognitive Dominance Workshop, sponsored by Lockheed Martin and Colgen, LP, was held at the WestPoint Military Academy in July 2006. The theme for this year was Assessing Intuitive Decision-Making Performance in Military Leaders. Three workgroups were formed to target the areas of assessment, training and technology, and ethics. The Ethics Workgroup and was led by Wendy Shaneyfelt. Patrick Becker from Colgen, LP documented the discussion (see Appendix D for a complete transcribe of the Meeting Minutes).

There were six participants in this workgroup representing military, private industry, and academia; US and Canada; female and male; and ranging in age from approximately 20 – 50 years old. Excerpts from the Meeting Minutes highlighting areas of risks related to cognitive systems are provided below in Table 2.4-3.

Table 2.4-3. Highlights from Cognitive Dominance Ethics Workshop

HIGHLIGHTS FROM COGNITIVE DOMINANCE ETHICS WORKGROUP **Questions and Specific Concerns**
1. 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? • 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.

- Who will guard the guardians?
- 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?
- 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.

2. Misuse of a cognitive system application.

- 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.

3. Abuse (use for a purpose that was not "good") of a cognitive system application.

- 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?

4. Accidental use of a cognitive system application.

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

5. 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.

- 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.
- 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?
- Are there health risks involved due to invasive enhancement processes? Infection, natural cognitive deterioration due to dependency, mental, emotional, psychological?
- 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.)
- Will there be universal application available to all? If so, is it effective for all?
- Is it ethical to not offer this technology if we have it?

6. Safeguarding cognitive profiles (privacy).

- Is a "consent form", "cognitive content disclaimer", or some other administrative process required?
- 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.
- Should there be a disposal plan or expiration date for cognitive models?

7. Cognitive system technologies for training could improve educational opportunities, skill sets, technical abilities, etc.

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

8. Cognitive system technologies for the purpose of selecting individuals via assessing and quantifying desirable traits.

- 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.
- Selection might preclude "out of the box" thinkers or other minority traits that add needed diversity and an element of unpredictability.
- Selection might help us better place individuals in more appropriate positions.
- Selection might be bias and produce unfair assessments.
- Selection might stifle diversity.

- Selection might foster too much conformity thereby producing a higher degree of predictability.

9. Accepting and embracing cognitive system technologies.

- Ensure adequate safeguards are developed and put in place to protect individuals from abuse
- 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.
- 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.
- Applying lessons learned from introductions of other previous risky technologies might help mitigate similar problems.

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

- 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?
- Will there be an asymmetrical development in some elements of the populace?
- Will this cause a shift in power, an adjustment in social norms or stratification?
- Will those who are cognitively enhanced be viewed as less human or just a more capable human?
- Would more inexperienced people who are cognitively enhanced be accountable for higher expectations?
- Would that which was once considered unethical become ethical?
- Need to monitor the impacts on other individuals, groups, societies, etc.

11. How do we monitor the risks?

- 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.
- 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.
- Other forums – the press, legislature, and government regulatory activities which have a requirement to either keep the public informed or to safeguard the public.
- Industry – marketing and publicity can help ensure that the public is informed and willing to accept the technology.
- Watchdog organizations will be needed to validate and assess technology; inform; counter positions; and monitor progress.

2.4.1.2.3. The Future of Neurotechnology: Social, Political, Ethical, and Legal Issues Workshop

Sponsored by DARPA, this workshop was held at the Potomac Institute for Policy Studies in July 2006. Participants included representatives from DARPA, the Potomac Institute, Sandia, Johns Hopkins University, George Mason University, and the University of California at San Diego. The focus of this workshop was to determine what social, political, ethical, and legal implications will revolve around neurotechnologies in the future. Highlights from the roundtable discussion are presented below.

- Procedures can be developed and applied to technologies intentionally developed for beneficial use.
- Funding sources effect public perception.

- The development of neurotechnologies is occurring globally.
- The possibility is real for legal regulations to halt/hinder beneficial technologies.
- Would this technology one day be used as a lie detector? Would it be accessible within our own homes? How would a mental illness effect the results? If it could be used to detect self-deception how could it effect that person psychologically or emotionally?
- There is no universal perception defining ethics. Some see a technology as ethical while other individuals, groups, religions, cultures, or countries do not.
- How do you objectify augmented cognition? Is augmentation for one individual appropriate for another?
- Applications of interest: prison, financial analysts, expert traders, legal courts, addicts.
- How would this impact selection of political leaders? Could their moral and ethical standards be quantified and compared?
- Will early adopters of the technologies give them an advantage to dominate a business environment or market? Could their be negative financial impacts?
- Do we want to raise the bar on our personal baselines by augmenting our cognition? Can we afford it? Can we all afford it? Would a gap form between the "haves" and "have nots" or "won'ts" and "won't nots"?
- Regulations or funding of the technologies could influence the balance of power. How will the technologies be distributed? Commercial interest will likely drive development if a profit is to be made.
- When will the shift occur from asking the question "Is it ethical to use it?" to "Is it ethical not to use it?"
- What are the ethical considerations of introducing a new technology?
- Dual uses of technologies are difficult to comprehensively define when the technology does not yet exist. Moderate introductions of technologies allow for discovery of dual uses.

2.4.1.2.4. Comments Captured from Internet Website in Response to Business Week Article

Chris Forsythe, a cognitive psychologist from Sandia and one of the founders of the CS&T area, was interviewed for an article entitled "The Ghost in Your Machine" published in the August 25, 2003 issue of Business Week magazine. Shortly after this article appeared, discussion about this article surfaced on an Internet website called Slashdot. www.slashdot.org is a website known as "News for Nerds". Highlights from the discussion that took place on Slashdot about the article are presented below.

I predict that if/when such a technology becomes prevalent, it will greatly reduce the human ability to make decisions.

Take for example any simple video game, how about MahJong (the stack of tiles that you have to match pairs on to remove them).

If you play it without the computer's aid, you develop a good eye for it and can do quite well. However, if you constantly hit the 'help' or 'hint' button, you become dependant on it showing you the next move, and never develop the skill for yourself.

To put it in context with other situations:
How many of you need a calculator to find a 10, 15, or 20% tip amount? Worse, how many of you need a calculator to add that extra 3.25 onto your 21.75 bill? I admit, it takes a great bit of effort for me to add simple numbers in my head simply because I don't exercise that ability enough.

I agree about that. I'd like to hear more guts, less fluff. Though it is rather fascinating fluff.

But they do seem to have a new idea: attempt to model the cognitive process of your user, notice where the results of your model differ from the user's actual behavior, and use those differences to improve your model.

It's applying concepts of machine learning to a good problem in an interesting way.

I think many of the posters here confused things a bit, the "cognitive" part refers to the thought processes driving the behavior of the user. Despite the use of the hype-laden term "Synthetic human" (that was only their *initial* goal) it's not claiming to be AI, only an adaptive model of the user's behavior patterns.

Until we have a technical paper that describes their approach in detail and can be peer reviewed I will remain skeptical.

Exactly. And given that it is a government-funded DARPA project, I suspect that we won't be seeing that for a while. Although many DARPA projects ostensibly have commercialization as the ultimate goal, if something like this were to really succeed it's highly likely that it would be tagged as valuable and/or dangerous and classified.

I share your frustration totally (sometimes Word expands my selection to include the punctuation at the end of a word... wtf!?). However, when people say "I don't need any help from my computer!" I feel they aren't thinking it through -- your computer is ***always*** assisting you to some degree. This notion of "overzealous assistance" is all relative. My mom needs AOL in order to "see the Internet" (it's like fingernails on a chalkboard when she says that to me), but I find the level of "help" that AOL provides as frustrating and cumbersome.

Perhaps that's where this technology could be put to it's best use: ***correctly intuiting exactly how much assistance*** you may need. So, for example, Clippy v2.0 would ask you if you need help with that letter when you are a newbie, but scale back the assistance when it intuits that you don't need help.

What happens when the user is a sick, twisted and sadistic person. Will the computer adapt to that kind of user?

And to focus on another problem: if this thing learns about you behavior, don't you mind about your privacy?

2.4.2. Focus Groups

2.4.2.1. Focus Group Organizational Structure

Prior to establishing the focus groups it was deemed important to try and establish some diversity across the focus groups with a somewhat homogeneous population with the focus group. The key factor was determined to be experience with cognitive science/systems. The Baseline focus group was a somewhat random sample of participants from the Cognitive Systems Workshop in Santa Fe reported in Section 2.4.1.2.1. This group could certainly be considered to be more interested in the science area, but the data indicated the sample varied from those who were expert to those who were new to the area. Four other focus groups were identified:

1. Non-Technical: UNM professors who had no technical background in cognitive science/systems.

2. Public: non-university persons who were well-educated (at least bachelors degree), but had no association with SNL or UNM.

3. Technical: UNM (CHTM, MIND) and SNL personnel who were researchers in the field of cognitive science/systems.

4. Surety: SNL personnel who were specifically associated with one or more of the surety disciplines, such as safety, reliability, security/use control, surveillance, human factors, or quality.

2.4.2.2. Focus Group Interview Procedure

The Focus Group Interview procedure is illustrated in Table 2.4-4.

Table 2.4-4. Focus Group Interview Procedure

Prior to Focus Group Meeting
1. Identify potential participant pools for the four separate focus groups. 2. Contact participants, with brief context information, and a request to participate. 3. Send briefing materials to participants prior to holding the focus group.
At the Focus Group Meeting
1. Participants complete consent form. 2. Participants complete questionnaire survey. 3. Recording of meeting discussion session begins. 4. 'Ground rules' of discussion reiterated: a. Goal to identify the range of views, concerns, and opinions b. Therefore all opinions are welcome and valuable. 5. Team member gives brief context of Cognitive Systems. 6. General discussion – role of team members is to: a. Keep the discussion within the broad topic area. b. Respond briefly to specific questions as necessary to enable the discussion to proceed c. Facilitate discussion as necessary. 7. Thank the participants, cease recording.
After the Focus Group Meeting
1. Transcribe recording and surveys; ensure participant identification is not on transcriptions. 2. Within three days after the meeting, send e-mail to participants, giving thanks, and asking if there are any further comments they would like make. 3. Hold project team meeting/telecon within one week of focus group for discussion and preliminary analysis of findings. 4. Write summary report within ten days/two weeks of focus group.

2.4.2.3. Additional Surety Group Discussion

In addition to the usual Focus Group procedure, the Surety Focus Group was asked to consider the range of risks (their own as well as the perceptions of the other Groups) and identify potential Surety methods/techniques that might be applied to reduce the risk in the key areas identified. A preliminary table summarizing the perceived risk areas derived from the other Focus Groups was provided and potential surety methods/techniques to reduce highest concern risk areas was discussed.

3. SUMMARY OF RESEARCH RESULTS

This section provides a summary of the research results derived from the research data gathered during this study effort. The research data is provided in Appendix D.

3.1. Focus Group Demographics

Focus Group demographics (age, gender, work area, and cognitive technology experience level) are summarized in Figures 3.1-1, 3.1-2, 3.1-3, and 3.1-4. As can be seen by the distributions:

Age Profile: biased toward "middle age"

Gender Profile: biased toward "Male"

Daily Work Profile: equally distributed among Technology Management, Technology R&D, and non-Technical activities

Cognitive Science Experience: equally distributed "Not Expert" to "Very Expert"

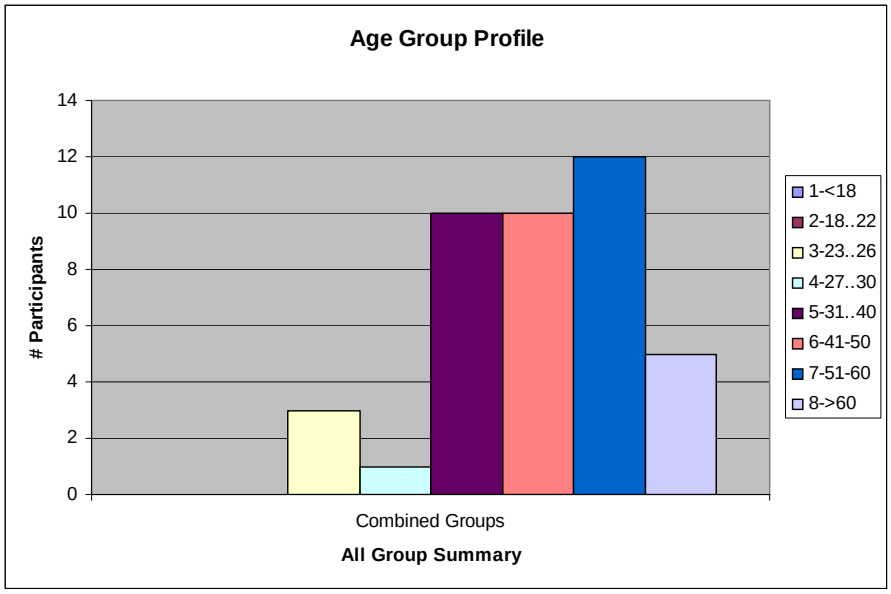

Figure 3.1-1. Focus Group Age Profile

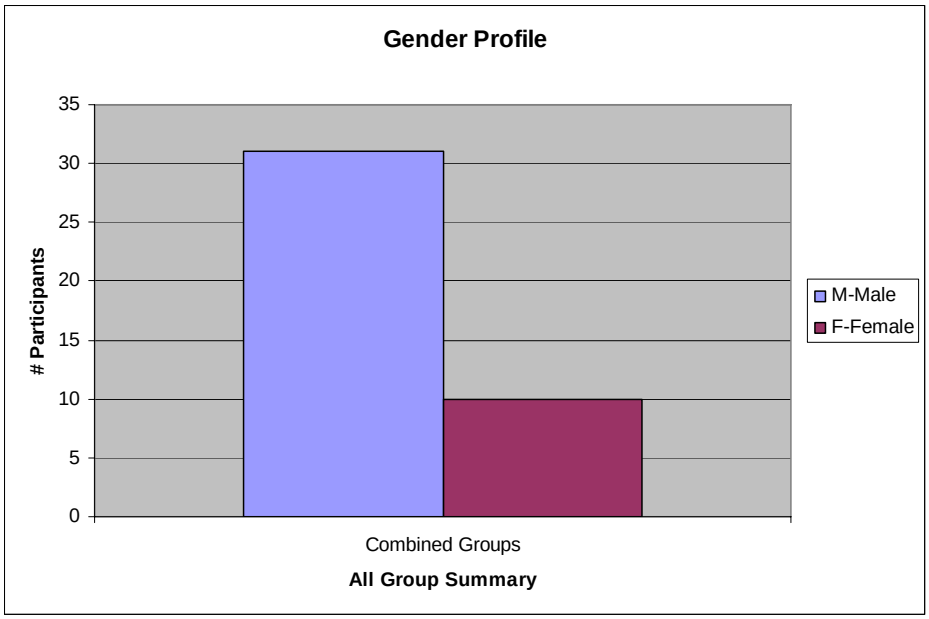

Figure 3.1-2. Focus Group Gender Profile

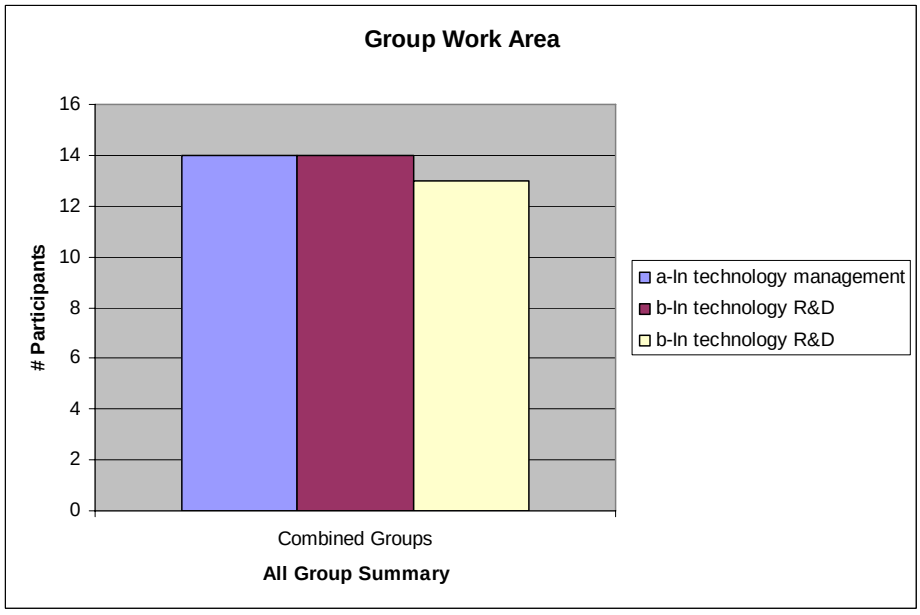

Figure 3.1-3. Focus Group Daily Activity Profile

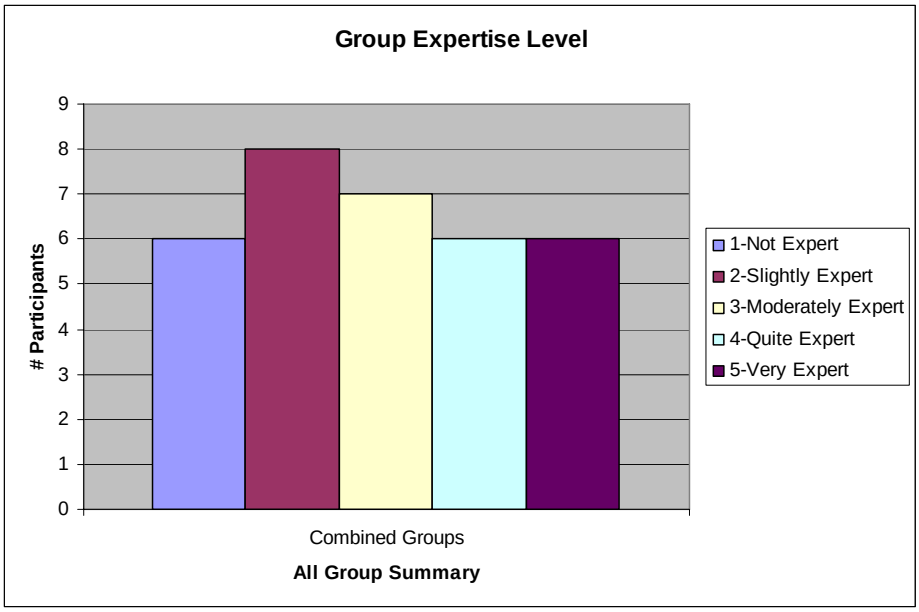

Figure 3.1-4. Focus Group Cognitive Technology Expertise Profile

3.2. Cognitive Technology Impact on Societal and Ethical Issues

The reference data, questionnaire survey responses, and focus group discussions provide a wide range of concerns (ethical and risk-based) related to the development, use, and/or sustainment of cognitive systems. The general response to whether cognitive technology would cause societal and ethical issues is illustrated in Figure 3.2-1 (by Focus Group) and Figure 3.2-2 (across all Focus Groups). These results illustrate:

Cognitive Technology Ethical Issue: there is a strong perception by each Focus Group and across all Focus Groups that it is likely cognitive systems would cause societal and ethical issues.

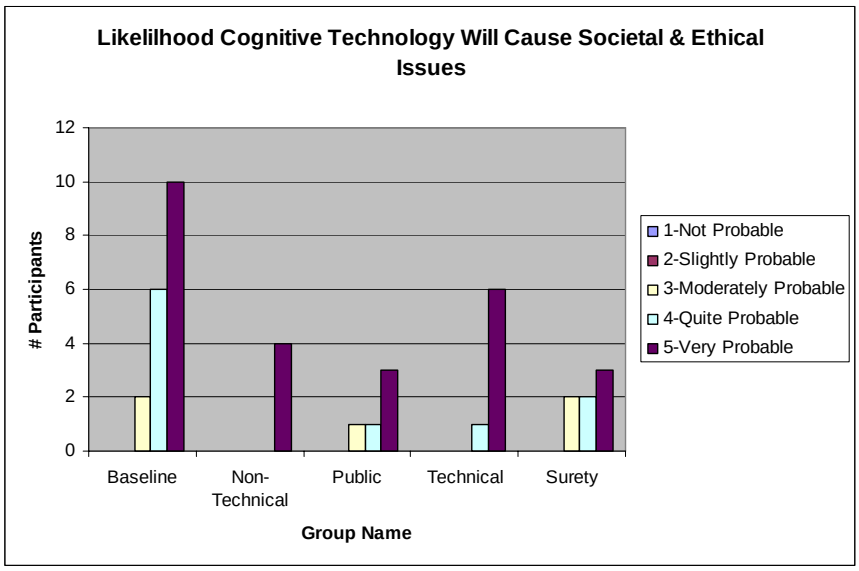

Figure 3.2-1. Cognitive Technology Likelihood to Cause Societal and Ethical Issues (By Focus Group)

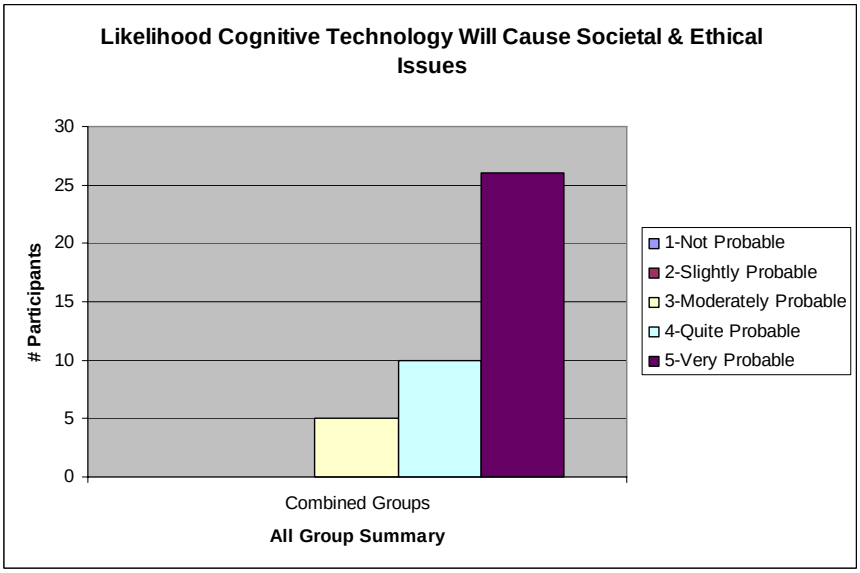

Figure 3.2-2. Cognitive Technology Likelihood to Cause Societal and Ethical Issues (All Focus Groups)

3.3. Application Area Risks

In all the Focus Group discussions, there was considerable confusion as to precisely what cognitive systems were and how they might be applied. Although the stated study focus of the cognitive systems was in the computational modeling of cognitive processes, the broader applications in physical, biological, natural, and technological sciences as well as many of the other risk dimensions were discussed. Nine application areas were identified in the questionnaire survey. Respondents were asked whether cognitive systems posed ethical issues for these nine application areas. The results of the 41 responses across all Focus Groups is illustrated in Figure 3.3-1. These results indicate:

Application Area Ethical Issues: The application areas of Medicine, Privacy, Human Enhancement, Law & Policy, and Military are of highest concern, although few participants felt any area was not likely to pose ethical issues. Ethical issues drive risks in many of the risk dimensions and specifically the perceived risk spectrum.

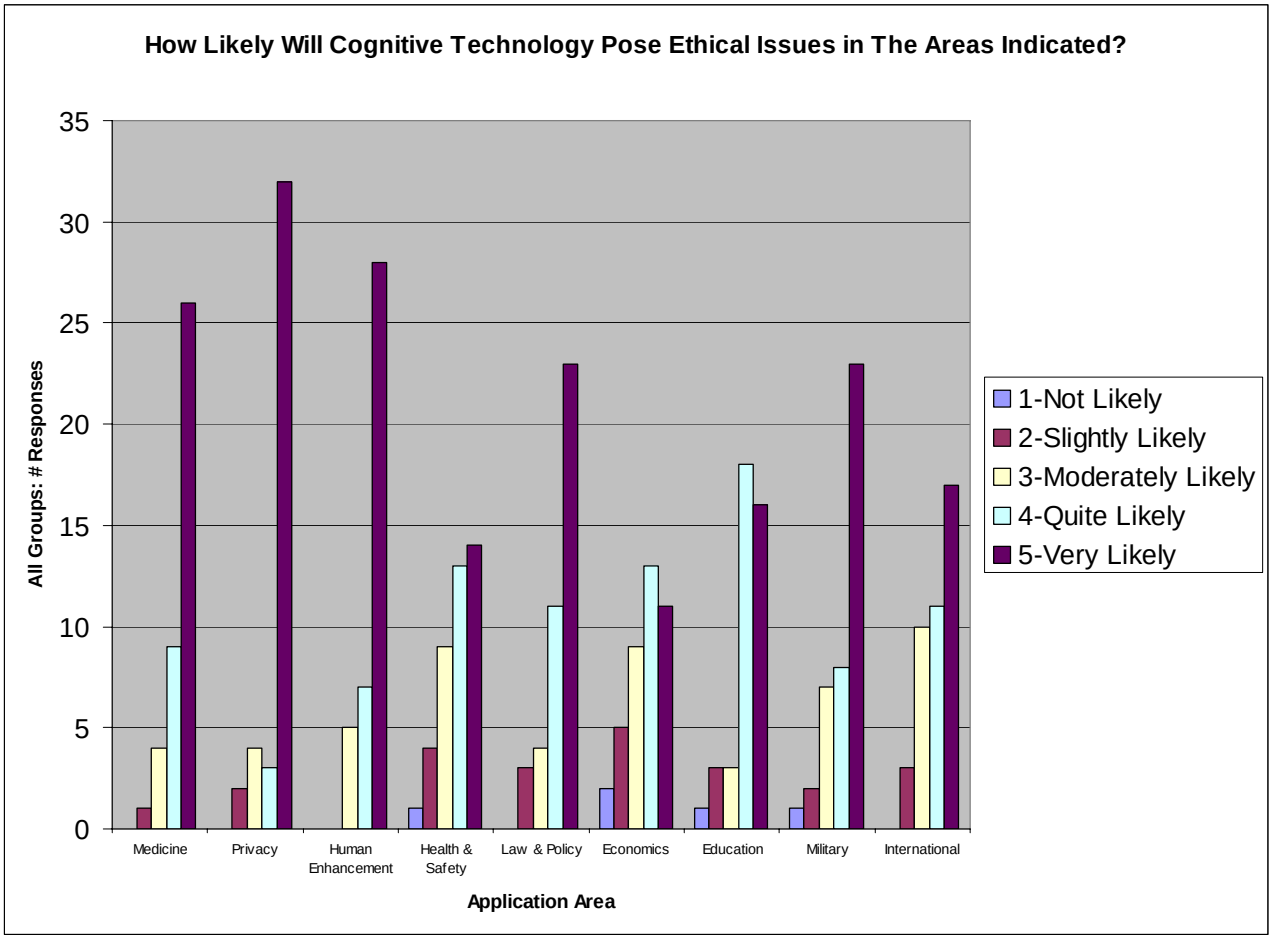

Figure 3.3-1. Summary of Perceived Application Area Ethical Risks (All Groups)

Some interesting comments from the Focus Group participants to these questions are indicated below.

> *"Privacy--Could one devise a mental signature that defines an individual? How can one control the capture of that model with the onslaught of technologies available to passively acquiring information about an individual? Education: Would the utilization of knowledge captures inhibit learning basic skill sets? Would someone lose out by not implementing/developing problem solving skills?"*
>
> *"Ethics seems hard to pinpoint. Seems to be based on religious beliefs as well as understanding of humanity"*
>
> *"The key ethical discussions will be how C.T. may be used against others and how it must be developed safely. C.T. also poses a threat to humanity if we ever develop truly sentient and autonomous machines, as they will easily outsmart us w/in one "Moore's Law" generation after that."*
>
> *"The extent of the ethical issues will likely depend on a number of facts, including whether we're enhancing cognition to return them to "baseline" following damage or pushing them beyond baseline. It will also depend on the degree of competitive advantage that enhancement offers. The question of 'mucking with' identity is another potential legal quagmire."*
>
> *"There are always costs associated with any advancement, however it seems that more focus should be put on the balance of the potential benefits vs. cost (ethical cost or otherwise). It's also not clear to me that ethics remain constant making it also difficult to make a decisive comment on."*
>
> *"If cognitive augmentation becomes a reality then it will become mandatory in the military and highly sought after in the private sector. Will it be available to someone who is not rich or a soldier/spy/etc?"*
>
> *"Issues of personality development could influence familial and personal relationships, particularly issues of guardianship."*
>
> *"Added Liability (Q4j) to list of areas for ethics concern. Validation is a key issue with cognitive models. Lack of validation will likely lead to a number of problems resulting in public distrust."*

3.4. Scientist Responsibilities

One area of potential concern that the questionnaire survey addressed was the responsibility of the cognitive scientists for the eventual societal results (both intended and unintended) of their research. Also, the question was posed as to how easy it would be for the cognitive scientists to recognize potential ethical issues associated with their projects. A set of questions was also posed as to the responsibility of cognitive scientists to communicate information concerning their projects with the public: explain to the public, discuss with the public, and consider public opinion in regard to the research direction. The survey results are illustrated in Figure 3.4-1 (overall responsibility), Figure 3.4-2 (recognition of ethical issues), and Figures 3.4-3 (explain), 3.4-4 (discuss), 3.4-5 (consider public opinion). Although the overall Focus Group results were fairly consistent, there were some outliers in the individual Focus Group results. The overall results indicate:

Overall Societal Responsibility: strong perception that the cognitive scientist is responsible for societal results (intended and unintended). Minority felt not responsible.

Ease of Recognizing Ethical Issues: there is a fairly equal distribution from "Not Easy" to "Very Easy" concerning the question whether it would be easy for a cognitive scientist to recognize ethical issues in there research. Disagreement may be due to perception of issues being application-dependent, and also to varying degrees of sensitivity among individuals towards potential issues.

Responsibility to Public: strong agreement that the cognitive scientists are responsible to explain and discuss their research with the public. Less agreement on whether the cognitive scientists should be responsible for considering public opinion when deciding the future direction of their research.

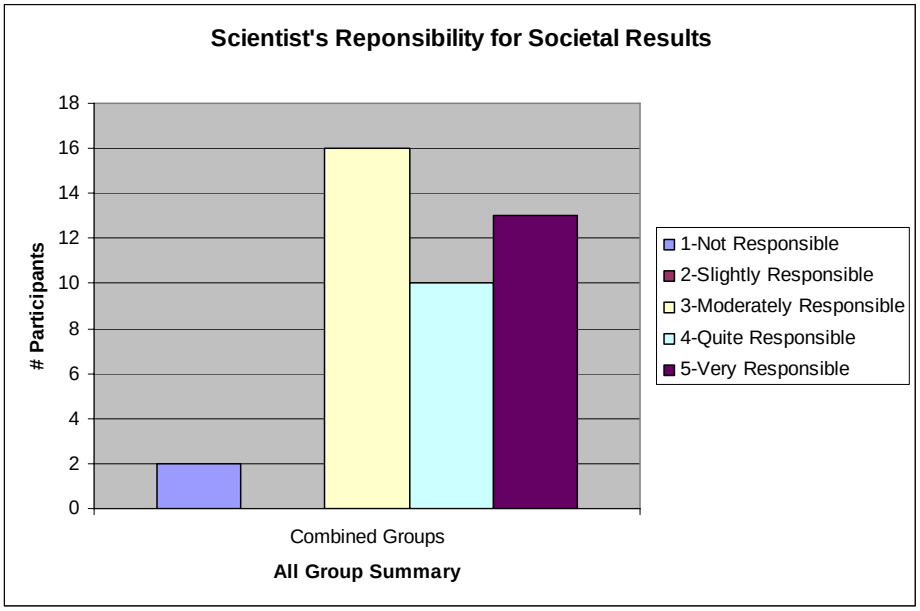

Figure 3.4-1. Cognitive Scientist Responsibility for Societal Results of Research (All Groups)

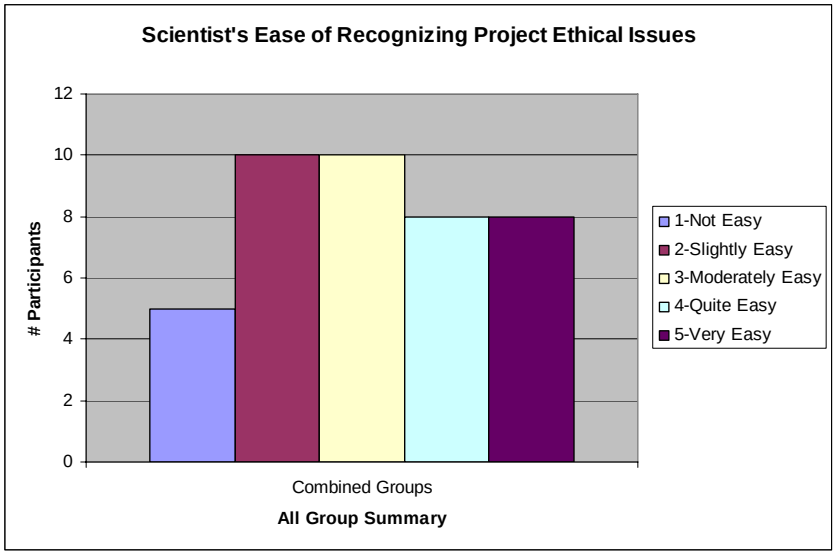

Figure 3.4-2. Ease of Cognitive Scientist Recognizing Ethical Issues (All Groups)

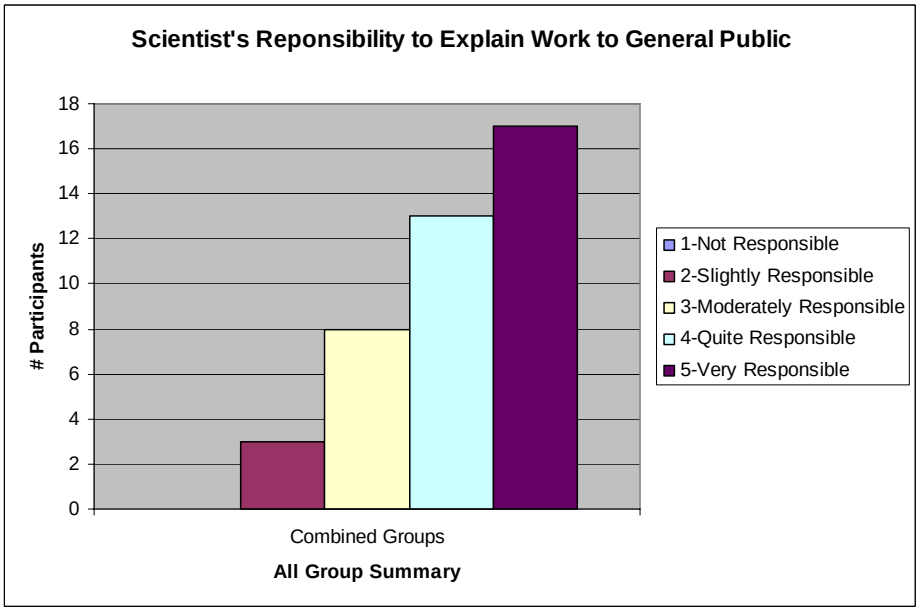

Figure 3.4-3. Cognitive Scientist Responsibility to Public: Explain Research (All Groups)

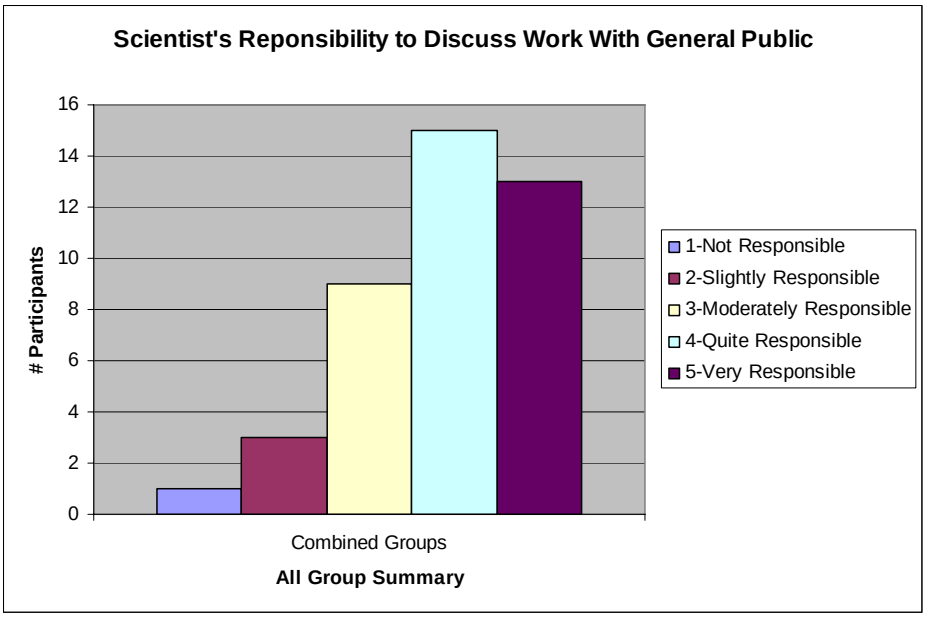

Figure 3.4-4. Cognitive Scientist Responsibility to Public: Discuss Research (All Groups)

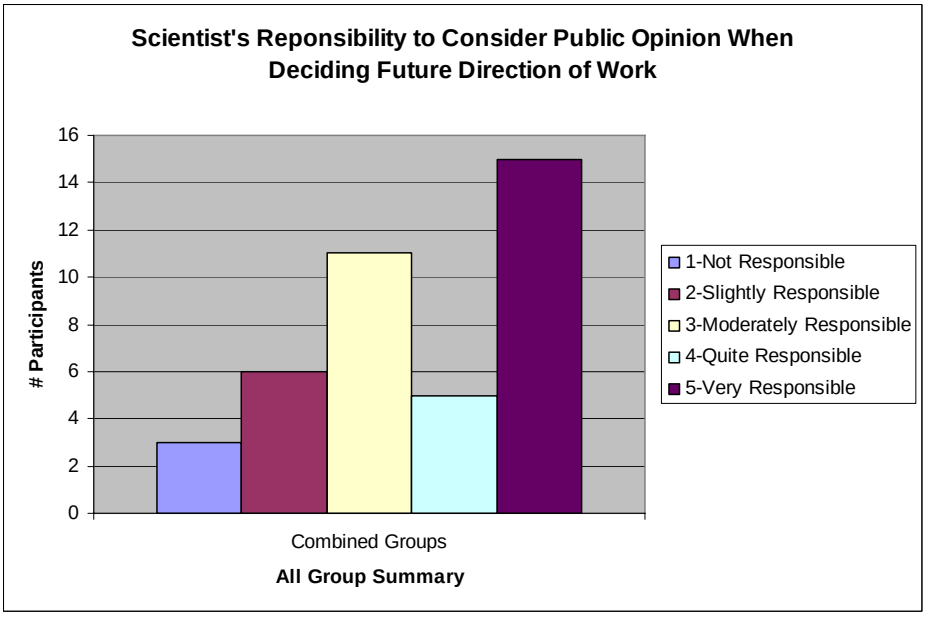

Figure 3.4-5. Cognitive Scientist Responsibility to Public: Consider Public Opinion (All Groups)

Some interesting comments from the Focus Group participants to these questions are indicated below.

> *"Researchers have to be honest about the claims for the results of their work."*
>
> *"There is no danger in knowing the truth, only in how it is applied. This is the responsibility of society as a whole, not just scientists. However, scientists should help inform."*
>
> *"Almost all technology can be used in many ways. Not doing something because it might be misused would be unfortunate. It's often tough to see the downside of something you're excited about. Educating the public--at least in a general way--about S&T is vital. But often the degree of expertise required to truly understand something is beyond the reach of the average Joe--including me. If we never moved beyond what the public was comfortable with, we'd all still be living in caves."*
>
> *"This is an interesting question. In a democracy system, I would put more emphasis on the entirety of society having responsibility."*
>
> *"Public opinion should be considered, but shouldn't restrict R&D."*
>
> *"Scientists develop technology--society and their leaders decide whether to use or not."*
>
> *"My experience is that as forthcoming as one is with the media--the story gets distorted by the time it gets printed. Thus I can understand scientists reluctance to 'talk to the public'."*
>
> *"Public opinion is a difficult and ill-defined term. If public opinion means debates and concerns in the public sphere, then my answer would be 'very responsible'. If it means a popular opinion or perspective, then I would say the scientist must be informed of it but not necessarily directed by it."*
>
> *"It may be possible for unfair/hostile gossip/reporting to harm scientific progress in this area. How can this be mitigated?"*

Further discussion of the ethical principles and guidelines for the development of cognitive systems has been investigated in [SHANEY-2006a].

3.5. Focus Group Discussion Highlights

Each of the four primary Focus Groups participated in very interesting discussion sessions of approximately one hour each on the topic of cognitive systems, ethical issues surrounding such systems, and potential benefits and risks such systems and the associated cognitive technologies might create. Files of the full transcripts (with no name attribution) of each Focus Group session as well as a summary highlight across the sessions are included in Appendix D. This section provides some of the more interesting comments across all the sessions. These discussion points, the questionnaire survey responses, and the LDRD research team analysis leads to the spectrum of risks and the potential use of surety methods to mitigate the risks that are summarized in Section 3.6 and Section 3.7.

3.5.1. Reliability

Reliability: *You'd monitor it and you'd mentor and coach it, and do all the things to try to get it up to the point where it would do things like you want them done. But what happens if you get into a risk or situation, for whatever reason, and you let your cognitive system take control, and it does something that causes harm to others, or whatever.*

Reliability: *What I'm trying to say is these benefits come with associated costs. We are placing values on things like intelligence, memory and attention. But those are value judgments. You're potentially creating a huge house of cards that will come tumbling down when you realize the lack of the model's ability to take into account the complexity of the brain.*

Reliability: *One thing I was thinking of is if you have something that is an implant you have it powered up in some way. It seems to me that looking at it from a weapon design perspective, this is a really bad idea – to take a functioning systems and couple something else which has some other capability to it. … You are adding … a sneak path for the system it's being coupled into. So, it can basically, in a certain sense, break the brain locally. … Or it can actually work as intended, but actually cause damage to the brain locally. Now this may be a relatively low risk in cases where you have it up in basically the mass storage of the brain. But I think if this got anywhere near a center of emotions; it could lead to some real problems. … If you have another power source in there that can potentially inject energy in ways that were not intended, it seems to me that anything which was sufficiently powerful to be interesting would almost, by definition, be impossible to rigorously screen for unintended functionality.*

Reliability: *Maybe you have behavior you don't really expect to get. And yes, those are concerns, absolutely. It's not like you don't know what you'd like to do – it's just whether you can do that safely and somewhat predictably. That's a real good question.*

Reliability: *What about verification and validation. Is that an issue you would see with these systems since they learn and are adaptive and complex? … Where are you supposed to validate? … If it's supposed to provide better judgments than you yourself provide, how do you validate that?*

Human Factors/Reliability: *So how about in the area of human factors? Is there anything in that area that you think about? … I could imagine that if you came up with a general cognitive model of how people behave in different context, you would be able to come up with situations in which people would be more likely to make errors. And then you could*

mitigate against those types of contexts. … It's a mistake-proofing kind of idea where we're saying 'what kind of things would you do to make sure you (do this?) only one way', and that's what you do. … So you can engineer those out rather than necessarily augment. … You could engineer out some of those kinds of environments if you anticipate where the error might occur.

3.5.2. Privacy

Privacy: *What about the risks in short term modeling an individual at some point? …. I can imagine it could be an ethical issue if the person who is modeled is not aware of being modeled. If you sold their model of them, they could say, 'who owns that model – the person who developed the software, or the person themselves?*

Privacy/Ownership: *To me, in our capitalist society, the property, who owns the intellectual and productive property of these things; how you negotiate that which should be negotiated. These are huge issues. You couldn't possibly see upfront when you allow this thing to model you in one way or another, even if it's not for you. If I let my model be your personal slave, do I have ownership over that? The level of thought processes that this thing can do may be at more contemporary levels, I would want ownership to continue over that. I would want a lot of money upfront, and a percentage of all future productivity of that thing. And privacy rights of that thing. …. Privacy. Absolutely.*

Privacy/Ownership: *One last thought – it's beyond where you're going, but we did work on electronic medical records four or five years ago. We brought doctors and patients and insurance companies. We asked what people were comfortable with. The variable we left out was who had the information. And it turned out to be the crucial variable. It turned out that people were very comfortable with their doctor seeing their medical records. In fact, they were very comfortable with any doctor seeing their medical records in an emergency. Then when you said 'an insurance company', they got really panicked because there was money involved. And when you said 'an employer', they got really panicked. And when you said, 'the government', they flipped out. It's the same thing with a model of me – my personality. … So an important factor isn't just 'what is the technology'; it's 'who has control?' … Ownership is clearly a risk. Who owns whatever the technology is, is clearly a risk. Who has access to that technology or information is clearly a risk.*

Privacy/Ownership: *A class of those risks that I want to bring up … is this idea of having something observe you either physically, internally, behaviorally – and creating some form of model of you for your own use in some way. But what about ownership and privacy and those [models] being stolen. What are the issues associated with that?*

Privacy/Ownership: *So given that you do have model created however, who owns it? … I worry about my privacy and who owns it. If I want to interview a whole new group of people to work in my group, maybe I dial up these models and review them and get some indication of how everybody thinks; who is technical and who is otherwise. The loss of privacy worries me. … It takes identify theft to a whole new level. Not only do they have your social security number; now they have your cognitive profile.*

3.5.3. Liability

Legal/Liability: *Well, from what I heard from an attorney was that there was no legal definition of 'human'. That, to me, raises an issue. At what point does whatever in my*

computer think it's human? And how can I tell it it's not, and all the ramifications thereof. To me, that probably indicates legislation down the road to define what's human. It's ultimately going to be needed.

Liability: *Let's get the lawyers involved in this. When you sue somebody because your thing went bad and you're . . . It said that was a burglar when it was really your brother . . . If you have a surrogate or a model, and it's used in some way that actually does something illegal or criminal or questionable, who is at fault? … It thinks my house is being broken into, so it makes the decision to lock all the doors, and there was a fire, and the kids couldn't get out. I could see all kinds of liability issues. … Liability and responsibility. … But legal consequences are certainly a risk in some way. … They may be a great benefit, too. … I could imagine that would keep you in check. … Somewhat, yes, if they were worried that they could be sued if their system was used inappropriately in some way. The governments might not use it that way – they may use them differently.*

3.5.4. Legal/Ownership/Intellectual Property

Ownership: *So what if the cognitive system fell in the wrong hands?*

Ownership: *There is that aspect of getting it in the wrong hands, and how do you protect it. It doesn't matter if it's my personal assistant. If that information gets into the wrong hands.*

Ownership/Identify: *That comes back to (…) the point about property in a sense of identity. If we're talking about that, then we're into a whole bunch of difficult questions like whether such a thing could even be owned anyway.*

Ownership/Moral/Privacy: *The capacity of that cognitive system to make moral decisions should be responsive to my private self as well as to my public self and my commercial self. That's what should make responsible ethical decisions. And if this cognitive system is purely geared to the commercial space or war space, or the public space, it won't have the dimensions it needs to make ethical decisions. Nor would it, if it were purely geared toward domestic space, be able to make appropriate decisions. The property implications connected to this are enormous the moment you start thinking about the ethical dimensions. Because the full ethical decision-making we want to have happen is something that humans are capable of because they work within a range of fields, and not a single field. … Right. Something that has moral authority – a morality to it – cannot be owned. We couldn't have a child and have someone own that child.*

Ownership/Privacy: *…who owns the intellectual and productive property of these things; how you negotiate that which should be negotiated. These are huge issues. You couldn't possibly see upfront when you allow this thing to model you in one way or another, even if it's not for you. If I let my model be your personal slave, do I have ownership over that? The level of thought processes that this thing can do may be at more contemporary levels, I would want ownership to continue over that. I would want a lot of money upfront, and a percentage of all future productivity of that thing. And privacy rights of that thing. …. Privacy. Absolutely.*

Ownership: *… raised the question about who controls this. My response was, 'of course, I decide whether we're going to copy the system of myself'. But I want to interrogate that assumption. An old model of evolution was 'those things that reproduce themselves individually, and their genetic [inaudible] gets passed [inaudible]. Current thinking about*

evolution isn't that way, and non-western cultures certainly aren't that way. Another model would be, 'your community decides whether to authorize you to reproduce yourself'. It's a collective decision about who ought to be cognitively modeled and who shouldn't be. It would gain a huge amount of attention with our western individualist bias, but it's another approach to the world. And in fact, current models of evolution, co-evolutionary thinking about cultural end, biological evolution, suggests its both the reproduction of individual DNA and the cultural selection for certain. Those things work in tandem, which would suggest you have to authorize the cognitive system of yourself, and your community has to say, 'yes, it's okay to copy you'.

Ownership/National Security: *We're all fallible in how we do things. So the question is, who is really doing this judgment here? Who is saying what these characteristics are? This is part of the ownership and who has control of saying who is good and who is bad? … Somehow there are principles behind how you do that. What are those principles? What exactly is it that we say, 'this is the way that makes it right'. Who is going to decide that? We here in the U.S. could say, 'we've decided we're the good guys, and this is the way we do things'. And somebody in some other country could say, 'that's not how we think about this'.*

Ownership: *If you worry about ownership and we could somehow protect that with your own private encryption where nobody could break it, that might give you confidence. … A biological key or something. … Ownership is also a threat. … If somebody else owns it.*

3.5.5. National Security

National Security: *The problem I see is that it's already being developed. Other nations are probably developing it. They may be ahead of us. It's going to continue. Strategically looking at the United States and what you're doing, how do you not continue development when you know other countries are going to develop it, and we better know what to do and how to deal with it. It's going to happen. We can't shut our eyes and say, "I don't want to deal with it because of these issues'.*

National Security: *If there are ways I can enhance myself that aren't really easily detectable, how are you going to stop me? And now that I'm doing it, you're going to do it, and now you're complicit. There's a lot that, even if the courts make a decision of what's acceptable, it may or may not be enforceable. … The previous revolutions like computers and biology, generally required a fairly large infrastructure and investment. This could be one where that direct investment is not required. If the United States, for some reason, would legislate against it, that doesn't mean anyone else is going to stop. Then if it turned out that it was a disruptive kind of technology that could be used against us, that's a scary scenario. … So what is the risk of the misuse of all this technology – no matter who it is that's misusing it.? The question isn't whether it will be developed. The problem is what do you do with it? How do you make it ethical? What do you do to limit its use in correct ways? Do we have ways to try to do that?*

3.5.6. Hype and Backlash

Hype/Backlash: *Along that line, I find it really curious that, if I were to use the words 'education' or training', even at the university level, this would probably perceived by almost everyone to be a good thing. But if I were to use the word 'cognitively enhanced', that conjures up things that might be 'bad' things. I'm not sure I know what the difference is.*

What keeps me awake at night is the possibility that these kinds of issues could, either maliciously or inadvertently, get elevated to a level of public attention that could prematurely end this debate or cut off the technology unfairly. So now the debate spins out of control before we even know what it is.

Hype/Backlash: *What does backlash and hype mean? … One of the concerns was that if we don't manage how we represent cognitive systems, there could be a social or legislative backlash to the technology that limits us. … It's like nuclear resonating. Used to call it Nuclear Magnetic Resonating; now we call it MRI. … The terminology sometimes blocks us completely. … If it's not packaged and played out properly, people could turn against it.*

3.5.7. Dependency

Dependency: *The technology almost invariably creates dependencies that are unsustainable. If the electricity goes off, if the water stops flowing, or if the toilet stops flushing, then what? And if people are too dependent on these … systems, then the proverbial … hits the fan, it's going to create a huge chaotic mess if people don't have a fundamental appreciation and understanding of their place in the environment and in the world.*

Dependency: *The assumption is that more technology is always better. It's actually a tradeoff. It creates more dependency. You may have special [capabilities] still being able to fly a plane and pay attention for 20 hours straight. But there is a tradeoff, like maybe you don't get along with your spouse as well when you get off duty or something. There is going to be some sort of cost to that benefit invariably. Everyone is looking at computation systems as this one directional way towards better understanding and capabilities, but what is the tradeoff of that? That's the huge blind spot that comes from not understanding how the brain works as a system. When you enhance one system, what are you going to be taking away from? …T here is an equilibrium that will be maintained at some level. It's a cost benefit analysis, and everyone is looking at the benefits without the costs, I think.*

Dependency: *What I am afraid of is we already have systems that are too complex for people to make the decisions about that. … So we have the power created that we might trust the computer more than the person. … So we're making systems that are so complicated, that only a machine can track them. But we don't have machines with wisdom yet. What are we going to do as we become dependent on systems that are too complicated for us?*

Dependency: *There's also an issue of atrophy. If you got sick, you rely upon something to do to enhance you all the time. … If something enhanced me continuously, now the batteries go out and it doesn't work, can I do it anymore?*

3.5.8. Diversity

Diversity/Ownership: *And who gets access to it also? Are we going to create a class of people who have access to these kinds of really high technologies who can enhance their life and go farther and farther ahead. Then the rest of the people will be left behind without the access, the ability, the money or the social structure where they will be able to use this. What is going to happen to all those people who don't have it? We would have to create schools. It increases the class differences in society. I see it moving it in that direction really fast already. There is a huge difference between people who have computers and those that don't. And if you have somebody who has this type of technology, it going to enhance all*

their abilities to do everything. It seems to me that it will increase the gap so much, we'll never be able to fix it.

Diversity: *I think there's an issue of losing your identify. I think that (…) could hire ten different people to do the job I do today, but there is something about me that uniquely can do it, there's an aspect of me that makes it special. And having something else that tried to do that as well would make me feel less valuable – that I could be replaced more easily.*

Diversity: *I think what separates this type of technology and its risk from others is that it speaks to the fundamental part of who we think we are, if it's done to the extent of that they are talking about. By definition, I think most people will find that to be a great personal risk.*

Diversity: *What about the very basic one – survival? … Survival of humanity? … Survival of the individual? … Identify self? Would that be in this same category? … You can lose your identify as long as you don't lose the reason for being.*

3.5.9. Equity

Equity: *Does the power serve the poor people, or does it serve the rich people? Do we take care of the hospital, or do we take care of the fire department? Are we worried more about the climate? Or about spotted owls, or the economy? These are tradeoffs.*

Equity: *Countries that were poor possibly and couldn't afford this technology would be even further down. So I'm not sure that's a positive thing for the world. Nor would it necessarily be positive to make a level playing field and say, "everybody's going to be equally bright"…*

Equity: *And you kind of wonder about long-term effects, too. That could be another risk if you took these enhancement drugs, how does it stress the system more? And once you're off that, how does it change how you age. … So if it could enhance your memory, how would it effect education and who could afford it? What would be those issues?*

Equity: *The question is if everybody took it, and everyone elevated their intelligence in equal amounts, would there really be an advantage to the world? Maybe that's a silly answer, but there are people who are probably below normal and can't function. They would benefit from going from non-functional to functional. But for everyone to get significantly brighter (inaudible), I'm no sure what that would look like.*

Equity: *I am unconvinced that cognitive systems could ever get to that level of what the Greeks call 'prudential wisdom' – distinguishing what situation I'm in, and therefore, which set of rules to apply. That's what raises a whole list of ethical issue for me. I totally believe in 100 years from now, we will be able to make these[decisions]. Can we make these in a way that we can trust them to make decisions that nobody has faced before? … It's the deep reaches of personality that may be out of reach of even the best cognitive systems.*

3.5.10. Human Enhancement

Human Enhancement: *One of the things I was looking at is it has huge public policy ramifications, and clearly the question that comes to my mind is when you're trying to mimic the brain and the functional aspects of that brain, it would appear that it raises some moral and ethical questions apropos. For example, what about the ability to use judgment? How do you discern that, and what brain do you pick? What comes from that? There are lot more questions because if we allow ourselves to be drawn into this software program that will*

think for us, or replicate for us where we get to the point of just cloning people. It raises a series of questions. I think those are some of the things that we would have to be cognizant of. But it does raise these questions in terms of the ability to adjudicate. ….. What about the individuality of the person?

Human Enhancement: *The thing that goes along with that is you're talking about cognitive systems and replication of the brain. How does it incorporate, or is it important to incorporate (I think it would be) feelings and emotions? They are individual. How do we develop that? Our feelings are each different. How does that play into decisions and judgments? How can you replicate that?*

Human Enhancement: *You've got possibilities of changing humanity indelibly in some way. We have all offline memory, and so we cease to have online memory. We've got those that aren't quite gray goo scenarios, but they are changing what it is to be human - kind of scenarios. Those are possibly – less likely. And then you have much more realistic inevitable scenarios like cognitive enhancements today – ways to all your students are no longer sleeping at critical times, and they're not on caffeine – they're on some pretty safe drugs. Do all the students get those drugs? What if I have offline memory – does every student get offline memory? What if I'm competing for a job, but I'm competing with cognitively-enhanced people. Why would anyone hire me? Then you bring up the steroid issues that exist in professional sports. But now they are there for cognitive jobs. If they level the playing field, do we all have to have it? Like 'Johnny' got his SAT scores, but they are artificially enhanced; 'Mike' should go to Harvard.*

Human Enhancement: *On the issue of people evolving. The scenario that comes to mind is when you think about training technologies where, in many trades, traditionally you have a system such that one has to go through extensive time and experience learning a trade and the nuances of how to something. And the respect that any individual receives within that trade is the basis of skills they've acquired – not just the skills they can demonstrate at any given time, but knowing the background that this person has 30 years of doing this. That is part of the respect that goes with the skilled tradesperson. When we imagine training technologies, we are in a very brief period of time. A person may not necessarily be provided the skills, but can be allowed to perform at a level equivalent to the skilled person. Not that they have the skills – they don't have the skills because they do not have the experience or understanding of all the nuances. Yet, they can perform at the same level. That completely changes the way we think about much of the maturation one goes through in their careers. It could significantly upset the balance of authority, esteem, and prestige.*

Human Enhancement: *There is a fundamental disconnection between what we know about brain function, and what we think we can model. … That disconnect is going to create blind spots. And those blind spots are going to manifest themselves in one of two ways: We are going to oversimplify brain functioning by thinking we can do more than we are actually capable of. Or, we are going to overlook some of the ethical implications because you're dealing with machines rather than humans. I think getting the computational and the neuroscience closer together somehow may decrease those blind spots substantially.*

Human Enhancement: *Do children need to learn addition, subtraction, multiplication and division anymore? Because calculators can do that. So that's a debate that is going on. …*

That's the nature of our world. We will need greater and greater technological assistance with any kind of important function that technology can assist us with. And we must adapt.

3.5.11. Moral/Religious/Spiritual

Moral/Religious: *To the extent that we're willing to assume what much of science assumes these days – humans beings as kind of what some scientists call the 'behaviorist model of human beings' - that we respond to stimuli. And this model can copy how I would respond to the stimuli in all kinds of incredibly impressive ways…But I don't think that is how we are as human beings. I think we respond much more fundamentally in ways driven by symbolic understandings of the world, and driven by passions, by desires, driven by aesthetic tastes. Not that they're complex, but that they're not rationally calculable that 'because I responded this way that time; therefore, I'm going to respond that way another time'. I think human beings are built in a different way. I can suspend that and say, 'you're going to build a model that actually responds in all those ways'. But the idea that this model is going to make the same "Art" that Participant 3 would make? I am morally, religiously and aesthetically committed to the notion of the human being that preserves a realm of freedom.*

Moral/Human Enhancement: *If this cognitive system truly has moral agency, presumably to create that, you have to so copy the human brain – that it also has the kind of fundamental lust for power that is present in human life, right? If it is truly a moral agent, and we don't control it, and it carries this lust for power, this is the kind of cyber future potential, right? It can be our tool. It can also be our dominator, right?*

Moral/Religious/Privacy: *So however the brain is functioning, if I screw up on something, it picks that up. For instance, if I were a sociopath, how would that be used in a legal system? What are the protections? What is the accessibility? What about the issue of confidentiality? … What ethical, moral, and religious implications are there?*

Morality: *Sometimes there are things like common sense or morality. How do you program those into a cognitive system?*

Moral/Religious/Spirituality: *The issue with spirituality and enhancement – in the discussions it's where do you draw the line? You can take antibiotics, though you weren't given those naturally. You wear glasses; you can have kidney transplant. At some point you have to say, in terms of spirituality, 'where do I stop?' And there seems to be something about the brain there. People start getting nervous about that.*

Moral/Religious: *(What about) the area of religion? … To me, some of this may be called spirituality, but in my case, I would object to some of the religious beliefs that God made me a certain way, and he wants me not to enhance it or to change it. That seems like a risk area that you would have to deal with when you have a technology solution to deal with it.*

3.5.12. Other Interesting Comments

Application Dependency: *I agree it would be a matter of application. What are you applying it to? For me, that's the essence of cognitive models. It's application. … So when you say you're talking about cognition, but you're measuring behavior, that's a pretty important disconnect for some applications.*

Application Variances and Consequence Priorities: *You've implicitly been referring to a broad spectrum of applications from those that are extremely invasive and high risk to those*

that are basically very … matter of fact, and may already be used -- like cochlear implants and stuff like that. It seems to me that we're missing an important point here. That is, it's very much like saying, 'well what's the difference between mass murdering and jay walking; they're both criminal activities. My argument would be that when things are quantitatively sufficiently different, they become qualitatively different as well. And where you make the cutoff and say, 'things to this side are qualitatively different then things to the other side is a matter for deep discussion. But I think it is plain that such a division can and has to be made.

Cognitive Systems?: *There is one problem associated generally with trying to integrate the ethical considerations about cognitive technology. And that is very often, you don't have a sense of what you're dealing with. One of the difficulties … is really understanding what cognitive systems are – where you bound that box. … Can you draw a line around understanding what you mean when you say 'cognitive systems'? ... I've noticed that some of us here are mentioning cognitive systems. But we have something different in mind. That might be a good idea to figure out what we mean here. There are several different major aspects here of cognitive systems. … I conclude that the term 'cognitive' is one that has been used so often and so broadly, it's become meaningless – there is no definition for the term.*

Cognitive Systems - Epistemology is Needed: *We have to articulate what it is to make a responsible human decision. And that will come out of establishing an epistemology. And that will be based on neuro-physiological understanding of what we are as humans. … I think we've got a lot of basic research to do to get into the foundations of what it is to be human. … [Need] an epistemology that supports [making] decisions in terms of taking responsibility.*

Cognitive Systems - Risk/Threats/Vulnerabilities Model is Needed: *Do you have a kind of threat model? … Risks only make sense in the presence of a threat model and the threat of some asset or set of assets that are important or that you're trying to protect, whether it's my ownership of medical records or something else. Well, that's my asset and I'm trying to protect it. What is the threat model? The threat model is that somebody that is unauthorized or misuses that information will have access to it. I think those are foundational elements that have to be defined in this research before going forward…. But without a clear definition of what's the benefit; what are the assets you're trying to protect; what is the threat model, and what are the potential vulnerabilities, then it just becomes a philosophical conversation. … But this is a very good idea of risk, threat, vulnerability, and looking at impact consequence. … If we're in certain domains where we say, 'the threat is not too bad', the asset is important, but still we could lose it and not be too concerned. The misuse couldn't be too bad because the threat couldn't take too much advantage of whatever it is, because the asset is not all that terribly important. We're not so concerned about that kind of situation. But when it starts to be our own whole behavior, suddenly it looks like that consequence could be a little higher than I'm willing to think about right now.*

Cognitive Systems - Surety Method Applications for Reliability/Protection: *I like the idea of encryption and access protection, and the idea of command disablement. If somebody gets a hold of it, and does something, it's going to disable this whole thing. So you can't use this. … If you have a device whose purpose was to surveil your mental state, then you could say, 'well I'm just going to use a different form of energy so that there's no way it*

can talk directly to the brain'. … But for this device to be useful, it actually has to have a way to convey energy and information into the brain. Some of these traditional safety methodologies really might not be applicable. … If you have the idea of power, then by current limiting what you trying to do is to say, 'I'm only going to allow a certain amount of power to get within certain regions.

Cognitive Systems - Technology Principles are Needed: *I have a sense that there are some fundamental principles here that need to be articulated and written down that people haven't thought of yet. … And we don't know quite how to do that.*

Cognitive Systems - Terminology is Needed: *If we could get to the point that if it's not the right word, let's gets rid of it; if we want to use that word, can we define it? Define it as well as physics is defined. I'm sure the physicists and chemists will find a boundary that they can argue about. That's probably true for cognition, too.*

Ethical Issues: *There are core ethical issues. And with cognitive systems, we're just seeing a new manifestation of a set of core ethical issues.*

Ethical Responsibility: *How do we continue to show that we are aware of risks? That we in fact are trying to control those and trying to do the right kinds of things? And make sure that, in the public perception, it comes across correctly so we don't get misinterpreted. That's a very big responsibility.*

Ethics and Scientific Responsibility: *So the question is not really whether things are going to be done like that. Because actually things are being done like that. It not like you can say, 'well, we shouldn't do that', because things are coming along like that. So the question really is what do you do when things are going to be like that? What is our responsibility? What is the responsibility of people developing it? What is the responsibility of … scientists to try and lower the risk as much as possible of misuse?*

Positive Perspective: *It's a true cognitive system. It could make decisions for you. If you have Alzheimer's and you can't make decisions on your own, you want to keep this thing around – drag it around with you – so you could be independent and live on your own, because you can do things – you just can't remember what you are supposed to do. … That's a positive one. That's very good.*

Positive Perspective: *I think in the foreseeable future, in the medical field, I can see tremendous advantages of cognitive systems to handicap, paraplegics, deaf, blind. I have been diagnosed with micro degeneration. When can you get me one? I'll be by tomorrow!*

Positive Perspective: *Back to your question about how would a memory-enhancing technology [be useful]. I think that would be valuable to me. I always look for tools to try to improve my memory; little ways to remember names and things and facts. It's very helpful to me to remember more in my day-to-day synthesis and work environment.*

Positive Aspects: *I hadn't thought about (cognitive systems work), but I'm excited that we are talking about it. … I can't wait to be there when you roll out the first model.*

Positive Perspective: *Well despite everything that has been said, I am very optimistic. At the root of my own optimism is there is so much of advantage that will come from being the technology developer. But the real advantage is going to be the technology developer who can successfully manage a risk.*

Positive Perspective: *I have a lot of confidence in the human being. I think the human is a very viable and logical thing. … I think in general, our human evolution is going to stay way ahead of our ability to model. Hopefully, that will be the check and balance on a lot of the things we do.*

3.6. Spectrum of Risks

The formal specification of what we mean by risk in this report is provided by the canonical Kaplan-Garrick risk triple [KAPLAN-1981]. This definition consists of three key components:

(1) What scenario can happen?

(2) What is the likelihood?

(3) What are the consequences?

The scenarios of interest are dependent on specific applications. Participants were clearly more concerned about some of the risk areas depending on which specific applications were being considered. For example, privacy was of considerable concern when the cognitive system involved modeling individual behavior, whereas of less concern when the cognitive system was modeling general group behavior.

Althaus [ALTHAUS-2005] reviews twelve dimensions in which risk can be analyzed: (1) linguistic and conceptual, (2) historical and narrative, (3) mathematical and logical, (4) scientific and measurable, (5) economic and decisional, (6) psychological and cognitive, (7) anthropological and cultural, (8) sociological and societal, (9) artistic and emotional, (10) philosophical and phenomenological, (11) legal and judicial, and (12) theological.

Within the above risk dimensions, following definitions are variously used in different risk literatures:

(1) Subjective risk: the mental state of an individual who experiences uncertainty or doubt or worry as to the outcome of a given event.

(2) Objective risk: the variation that occurs when actual losses differ from expected losses.

(3) Real risk: the combination of probability and negative consequence that exists in the real world.

(4) Observed risk: the measurement of that combination obtained by constructing a model of the real world.

(5) Perceived risk: the rough estimate of real risk made by an untrained member of the general public.

As described in [ALTHAUS-2005], embedded in these definitions is a specific distinction between risk defined as a reality that exists in its own right in the world (e.g., objective risk and real risk) and risk defined as a reality by virtue of a judgment made by a person or the application of some knowledge to uncertainty (e.g., subjective risk, observed risk, perceived risk). Whereas the former considers the metaphysical properties of risk, the latter is what can be termed an epistemological approach to risk. Taken as an epistemological reality, risk comes to exist by virtue of judgments made under conditions of uncertainty.

Clearly, the concepts of epistemological reality and metaphysical properties of risk are both of interest to the cognitive system risk spectrum. In considering surety methods that might be applicable to reduce specific risks relative to specific applications, we typically apply the metaphysical properties of risk and do not apply such a complex taxonomy of risk. Rather,

the surety methods typically are restricted to dimensions "3", "4", and "5". However, because cognitive systems are part of an emerging science, there is an abundance of "uncertainty", "unknowns", and judgments made under such conditions of uncertainty associated with their concept, development, implementation, use, and sustainment. Hence, it is worth keeping this complexity in mind to better understand the notion of risk aversion that is required when considering cognitive systems.

A preliminary spectrum of perceived risks is illustrated in Table 3.6-1. This table has significant overlap with Table 2.4-2 derived from the 2006 Cognitive Systems Workshop.

Table 3.6-1. Preliminary Perceived Risk Spectrum

Risk Area	Rationale/Description/Concern
Backlash and Hype	inaccurate reporting or self-serving bias about cognitive systems that creates barriers to conduct of legitimate research and applications
Dependency	we won't be able to get along without relying on the technology; what happens to individuality and creativity
Equity Issues	who has access to technology, gap between rich and poor, and so forth
Fallibility	reliability/uncertainty of cognitive system operation; how to develop, test, validate such systems for specific application use
Human Enhancement	long-term side effects; privileged use for personal advantage;
Identity/Self	what is the relationship between, say, my "self" and my cognitive model
Inevitability	these changes are coming whether we like it or not; also, inevitability that despite claims specific research/application will not be done, the bottom line is that it might be – either by one country or another
Liability/Legal	who is responsible for cognitive systems that malfunction or perform actions that result in catastrophic hazards; corporate protection of individual/group through laws and regulations
Loss of Diversity	one particular way of thinking becomes privileged because it is embedded in a widely used cognitive model
Moral/Religious/Faith-Based/Spirituality	conflicts between faith-based beliefs and scientific discovery; beliefs that such research is in conflict with faith-based beliefs
Ownership	who controls my cognitive system and who benefits from the associated intellectual property
Privacy	individual information protection; access control to cognitive model
Skepticism	What are the values that would govern an "ethical" system, and who would choose those values; skepticism in the capability to create cognitive models with any accuracy/predictability
Trade-Offs or Opportunity Costs	the sense that enhancing one kind of cognition may mean losing other kinds

3.7. Surety Methods and Applicable Areas

The Table 3.6-1 was discussed with the Surety Focus Group to determine potential application of surety methods. That table is updated in Table 3.7-1 based on the results of all focus group discussions and survey questionnaire responses, with the addition of suggested surety applications that would possibly be effective in reducing risk. For the most part, surety methods are applicable to the following risk dimensions [ALTHAUS-2005] (also, see Appendix B.2, Definitions):

(3) mathematical and logical

(4) scientific and measurable

(5) economic and decisional

The surety methods may be applicable to support arguments in the other risk dimensions, but are probably less effective as a direct capability to reduce risk. In addition, the surety scientists provided some thoughts on the prioritization of the risk areas indicated in Table 3.7-1. Privacy, ownership, and fallibility issues were by consensus of highest importance. In some instances there was a difference between personal and societal importance, although privacy, ownership, and fallibility were seen to be of high importance on both a personal and societal level.

Table 3.7-1. Perceived Risk Spectrum With 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 neuroscience, which currently lacks a broad theoretical basis, implies a high potential for unintended consequences.	Safety Principles Reliability: FMEA/FTA/PM/HF Methods/Sensitivity Analysis Risk Analysis: QMU Quality Methodology Ongoing monitoring efforts to detect adverse consequences early.
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 contribute to an increasing separation of humankind from the natural world. Will this reliance cause human abilities to atrophy?	Redundancy, system backups, and high reliability in systems will be crucial to 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	Emerging technologies are creating unprecedented possibilities for shaping

	irreversible and perhaps inappropriate changes to the course of human evolution.	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 wellbeing. 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.

4. CONCLUSIONS AND RECOMMENDATIONS

4.1. Key Conclusions

Conclusion 1: There was general consensus among the project Focus Groups that there are many areas of risk in the development and use of cognitive systems. In particular, privacy, legal/ownership/IP, and reliability of the systems seemed to be of highest priority. However, the risks and their priorities were clearly dependent on the intended application.

Conclusion 2: These high priority risk areas of privacy and reliability are particularly susceptible to surety analysis, through techniques such as encryption technology and reliability methodologies.

Conclusion 3: There is a strong perception by each Focus Group and across all Focus Groups that it is likely cognitive systems would cause sociological and ethical issues. The application areas of Medicine, Privacy, Human Enhancement, Law & Policy, and Military are of highest concern. Ethical issues drive risks in many of the risk dimensions and specifically the perceived risk spectrum.

Conclusion 4: There is a distribution of opinion as to how easily cognitive scientists would be able to recognize ethical issues in their research. This may be because the responses were given with various applications in mind, and/or to varying sensitivity to potential issues among respondents. There is strong agreement that cognitive scientists have a responsibility to explain and discuss their research with the public.

Conclusion 5: There is general agreement that cognitive scientists have a responsibility to consider the societal outcomes of their research. There is less agreement on whether the cognitive scientists should be responsible for considering public opinion when deciding the future direction of their research. Cognitive scientists participating in the project Focus Groups gave clear evidence of their concerns over a range of these issues. They also reflected a sense that other societal institutions, including the legislature and the judiciary, will play a major role in this process.

Conclusion 6: Surety methods provide assurance that a system is safe, reliable, secure, built with human factors considerations, and with an appropriate level of quality for the intended application use. Clearly, these surety methods are applicable to specific aspects of cognitive system design, and if appropriately applied would reduce the technical risks associated with such systems. Other risk dimensions, such as the appropriate distribution of the benefits and burdens of technology, for example, are less susceptible to resolution through surety methodology, and fall into the area of public policy and democratic decision-making. The development of reliable and safe technologies, however, will greatly facilitate this public policy debate.

Conclusion 7: The Focus Groups were designed as a pilot study to indicate the qualitative dimensions of the risks associated with cognitive systems. The groups broadly agreed on many of the queried issues, irrespective of group make-up. Future efforts will be designed with attention to diversity across age, gender, and work background.

4.2. Key Recommendations

A key observation of this study was the lack of a clear definitional framework within which cognitive systems could be clearly identified, categorized as to function, and associated with application areas. One needs an epistemology basis for cognitive systems, with a well-defined study of the theory, knowledge base, and reference to its limits and validity of application. A starting point would be to at least develop a cognitive systems epistemology framework within which existing research and gaps in information could be identified and systematically integrated. In order to adequately study the potential concepts of operation, operational scenarios, vulnerabilities, threats, and ultimately the potential for unwanted events (risks) associated with cognitive systems, it is essential to have such a framework. Some aspects of this framework are being investigated as part of [WAGNER-2006b].

Recommendation 1: It is recommended that the initial research results of this LDRD study be published in a recognized journal to acquaint the research committee with this effort.

Recommendation 2: It is recommended that a follow-on LDRD study be initiated to investigate and develop a cognitive systems epistemology framework, integrating within this framework the risk areas/issues, applicable risk dimensions, and surety methods identified in this preliminary study.

APPENDIX A - REFERENCES

[ALTHAUS-2005]	Althaus, C.E., "A Disciplinary Perspective on the Epistemological Status of Risk," Risk Analysis, Volume 25, Number 3, 567, 2005.
[AMA-2006]	American Medical Association, "*Principles of Medical Ethics,*" July 06, 2005. http://www.ama-assn.org/ama/pub/category/2512.html
[AMENDOLA-2001]	Amendola, A., "Recent Paradigms for Risk Informed Decision Making," Safety Science, Volume 40, 17–30, 2001.
[DG10100-2003]	DG10100/B, „The Process for Achieving Nuclear Weapon Safety at Sandia National Laboratories," Design Guide, Issue B, 2003.
[FORSY-2006]	Forsythe, C., Bernard, M., Goldsmith, T., "*Cognitive Systems: Human Cognitive Models in Systems Design,*" Lawrence Erlbaum Associates, Inc., 2006.
[GREELY-2005]	Greely, H. T., "*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
[KAPLAN-1981]	S. Kaplan, S. and Garrick, B. J., "On the Quantitative Definition of Risk," Risk Analysis, Volume 1, Number 1, 11–27, 1981.
[MCLEAN-2005]	McLean, Margaret R. "*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
[NAS-1995]	National Academy of Sciences, "*On Being a Scientist, Responsible Conduct in Research,*" National Academy Press, Washington, DC., 1995.
[NIH-2005]	National Institutes of Health, "*Regulation and Ethical Guidelines,*" Office of Human Subjects Research, December 2005. http://www.ohsr.od.nih.gov/guidelines/guidelines.html
[RESER-2006]	Memorandum: from Reser, Terry (SNL Human Studies Board (HSB) Administrator) Proposed "Investigating Surety Methodologies for Cognitive Systems" Study (SNL0652) is Exempt, July 10, 2006.
[SHANEY-2006a]	Shaneyfelt, W. L., "*Ethical Principles and Guidelines for the Development of Cognitive Systems,*" SAND2006-0608, Sandia National Laboratories Report, May 2006.
[TRUCANO-2006]	Pilch, M., Trucano, T., Helton, J., "Ideas Underlying Quantification of Margins and Uncertainties (QMU): A White Paper, " SAND2006-5001, September 2006.
[WAGNER-2006b]	"CS&T Plan (draft)," SAND2006-xxx, Sandia National Laboratories Report, October 2006.

APPENDIX B - GLOSSARY

B.1 Acronyms

CS&T	Cognitive Science and Technology
CTHM	Center for High Technology Materials
DARPA	Defense Advanced Research Projects Agency
FMEA	Failure Modes and Effects Analysis
FTA	Fault Tree Analysis
HF	Human Factors
MIND	Mental Illness and Neurosciences Discovery
NSF	National Science Foundation
ONR	Office of Naval Research
PM	Probabilistic Methods
PCMM	Predictive Capability Maturity Model
QMU	Quantification of Margins and Uncertainty
SNL	Sandia National Laboratories
UNM	University of New Mexico

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 System	Cognitive systems consists of technologies that utilize as an essential component(s) one or more plausible computational models of human cognitive processes.
Cognitive Technology	Products that aid a person's cognitive functioning (comprehension, 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.
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.
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.

	(10) 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.

APPENDIX C - PROJECT LEAD BIOGRAPHIES

C.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.

C.2 Wendy Shaneyfelt, SNL

WENDY SHANEYFELT is Computer Software Researcher at Sandia National Laboratories in the Cognitive and Exploratory Systems group. She received her Bachelors degree in Computer Engineering from the University of Nebraska. Her current work involves projects that detect patterns and anomalies in large data sets, automate knowledge capture, and develop and apply cognitive models of individuals. Wendy has developed ethical principles and guidelines for the development of cognitive systems and is structuring a plan to apply these mechanisms to real-world challenges. She has served as the leader for several ethics workshop groups to explore the societal, legal, ethical, and political implications surrounding cognitive systems technologies. Previous to her work in cognitive sciences, she worked as a software engineer for 20 years in the aerospace and defense industries.

C.3 Eva Caldera, JD, UNM

EVA CALDERA is the Associate Director, Institute for Ethics and a Research Professor of Law at the University of New Mexico. She received her JD from the Harvard Law School. She is engaged in teaching, research and community outreach in law, biomedical ethics and ethics of emerging technologies. Eva's courses include Bioethics (UNM School of Law, Spring 2005); Law and Bioethics (UNM University Honors Program, Fall 2005); Law and Science (UNM School of Law, planned for Fall 2006). She is a member of the Human Research Committee for UNM Health Sciences Center (appointed August 2005). She provides general consulting services to Sandia National Laboratories to review code of ethics for cognitive systems design.

C.4 Tom Caudell, PhD, UNM

TOM CAUDELL is Director of the Center for High Performance Computing Visualization Laboratory and Professor of Computer Engineering at the University of New Mexico. Tom received his PhD in Physics from the University of Arizona. His general area of research is in Computational Intelligence and Advanced Human-Computer Interfaces. Specifically, his research program addresses 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.

C.5 Kirsty Mills, PhD, UNM

KIRSTY MILLS is the Associate Director of the Center for High Technology Materials (CHTM) and Research Associate Professor in the Electrical and Computer Engineering Department at the University of New Mexico. She has a PhD from the University of Nottingham, England in Electrical Engineering. Her areas of current research interest include: Societal and Ethical Implications of Emerging Technologies; Nanotechnology Science Education and Outreach; and Human-Technology Interface.

APPENDIX D - RESEARCH DATA

The research data obtained during this study primarily consists of two parts: data from questionnaire surveys completed by the participants and discussion information from recordings captured during the various focus group meetings. In addition, normative reference/baseline survey data was collected from a workshop that served to prototype the questionnaire survey. Some interesting discussion material has also been obtained from several workshops/conferences associated with this effort. In all cases, names of participants have been removed from any of the information obtained during this study to keep the information anonymous.

Reviews were conducted at both SNL and UNM on the research approach for this study to ensure the use of human subjects in the study would be conducted in accordance with applicable regulations and requirements. In accordance with reference [RESER-2006], the following statement was made:

> *"Sandia's Human Studies Board has reviewed the information submitted for this proposed activity and has determined that the activity does constitute human subject research, but is Exempt from further HSB review in compliance with 10 CFR 745.101(b)"*

This determination was based on the stated scope of the proposed study and the description of proposed activity. The scope of work and proposed activity did not change during this project. The research team responsibilities, as well as those of the HSB, are described in detail in the HSB Procedures Manual (http://www.sandia.gov/health/hsb/HSBmanual.pdf).

D.1 Focus Group Interview Materials

D.1.1 Participant Consent Form and Survey Questionnaire

The following participant consent form and survey questionnaire were completed by each participant during the Focus Group Interview Sessions. In each case, the consent form was separated from the survey questionnaire so as to keep the information anonymous.

Consent Form and
Questionnaire Survey

D.1.2 Focus Group Interview Session Procedure

The following procedure was used during each of the four Focus Group Interview Sessions.

Focus Group
Interview Procedure

D.1.3 Pre-Meeting Participant Reading Material

The following read-ahead material was distributed to the focus group participants prior to each 90-minute discussion meeting. The purpose of this read-ahead material was to give those unfamiliar with cognitive systems some application examples to consider. They also served to facilitate discussion with those groups unfamiliar with cognitive science and technology.

Participant Reading Material

D.2 Workshop and Conference Discussion Data

D.2.1 Poster Presentation: Where Do You Draw the Line?

The 2006 Cognitive Systems Workshop: Bridging Cellular to Social was held in Santa Fe, NM on June 27-29. Eva Caldera and Wendy Shaneyfelt presented the poster "Where Do You Draw the Line?" The objectives of this poster presentation were 1) to stimulate discussion on the spectrum of risks associated with cognitive systems technologies, and 2) to elicit feedback from members of the cognitive systems community describing both real and perceived risks. To achieve these objectives, case scenarios were presented in the poster presentation with white space provided for informal comments. Surveys were also distributed to provide a more formal method of feedback. The full summary of the workshop results is in the embedded icon below.

Poster Presentation

D.2.2 Cognitive Dominance Workshop

The 2nd annual Cognitive Dominance Workshop, sponsored by Lockheed Martin and Colgen, LP, was held at the WestPoint Military Academy in July 2006. The theme for this year was Assessing Intuitive Decision-Making Performance in Military Leaders. Three workgroups were formed to target the areas of assessment, training and technology, and ethics. The Ethics Workgroup and was led by Wendy Shaneyfelt. Patrick Becker from Colgen, LP documented the discussion. There were six participants in this workgroup representing military, private industry, and academia; US and Canada; female and male; and ranging in age from approximately 20 – 50 years old. Excerpts from the Meeting Minutes highlighting areas of risks related to cognitive systems are provided below.

Cognitive Dominance Workshop

D.3 Focus Group Questionnaire Survey Data

The questionnaire survey data was entered within an Excel Spreadsheet. Graphs of the data were generated. This information is contained in the indicated file below.

Survey Analysis

D.4 Focus Group Discussion Recording Data

Each of the primary four Focus Group discussion sessions was recorded, and the recording transcribed as a record of the discourse. The transcribed information was carefully edited to ensure anonymous participant identification. Some difficulties with interpreting the precise audio discourse were experienced. A summary of the discussion data across the four sessions is included as well as transcripts from each of the four sessions. Identification of study authors was allowed in the session transcripts to distinguish comments that were not made by the Focus Group participants.

D.4.1 Summary Transcript Across All Sessions

Focus Group
Transcript Highlights

D.4.2 Non-Technical Focus Group Session Transcript

Focus Group
Transcript-Nontechnic

D.4.3 Public Focus Group Session Transcript

Focus Group
Transcript-Public

D.4.4 Technical Focus Group Session Transcript

Focus Group
Transcript-Technical

D.4.5 Surety Focus Group Session Transcript

Focus Group
Transcript-Surety

DISTRIBUTION

34 Sandia National Laboratories Internal Distribution

#	Mail Stop	Name, Organization
10	MS 0638	David Peercy, 12341
1	MS 0638	Nick DeReu, 12341
1	MS 1005	Russ Skocypec, 06640
5	MS 1188	Chris Forsythe, 06641
10	MS 1188	Wendy Shaneyfelt, 06641
3	MS 1188	John Wagner, 06641
2	MS 9018	Central Technical Files, 8944
2	MS 0899	Technical Library, 04536

15 University of New Mexico Distribution
ATTN: Kirsty Mills (5)
ATTN: Eva Caldera (5)
ATTN: Tom Caudell (5)
Center for High Technology Materials
MSC04 2710
1313 Goddard SE
Albuquerque, New Mexico 87106-4343