DEFENSE THREAT REDUCTION AGENCY
Scientific & Technical Review Information

PA CONTROL NUMBER:	PA 11-415 13 June	SUSPENSE:	8 Jul 2011
PM / PHONE / EMAIL:	Nicole Whealen 703.767.6354	DATE:	8 Jun 2011
BRANCH CHIEF / PHONE /EMAIL:		DATE:	
DIVISION CHIEF / PHONE:		DATE:	
DIRECTORATE / DIRECTOR /PHONE	Rob Gregg 703.767.5187	DATE:	
ENTERPRISE / OFFICE / PHONE:		DATE:	
PUBLIC AFFAIRS:	for RICHARD M. Cole (Chief, PA) / J Gaines	DATE:	6/28/11

1. TTLE: Challenges in Computational Social Modeling and Simulation for — CONTRACT NUMBER

ORIGINATOR McNamara, Lauren; Trucano, Timothy; Gieseler, Charles; Perry, Jennifer

2. TYPE OF MATERIAL: [X] PAPER [ ] PRESENTATION [ ] ABSTRACT [ ] OTHER

3. OVERALL CLASSIFICATION: [X] CONTRACTOR UNCLASS [X] PROJECT MANAGER UNCLASS

A. Review authority for unclassified material is the responsibility of the PM. Your signature indicates the material has undergone technical and security review.

B. Warning Notices/Caveats: [ ] RD [ ] FRD [ ] CNWDI [ ] NATO RELEASABLE [ ] SUBJECT TO EXPORT CONTROL LAWS

C. Distribution Statement:

[X] A. Approved for public release; distribution is unlimited (unclassified papers only).

[ ] B. Distribution authorized to U.S. Government agencies only; (check the following):

- [ ] Contractor Performance Evaluation
- [ ] Foreign Government Information
- [ ] Administrative or Operational Use
- [ ] Specific Authority
- [ ] Premature Dissemination
- [ ] Proprietary Information
- [ ] Test and Evaluation
- [ ] Software Documentation
- [ ] Critical Technology

[ ] C. Distribution authorized to U.S. Government agencies and their contractors: (check the following):

- [ ] Critical Technology
- [ ] Specific Authority
- [ ] Administrative or Operational Use
- [ ] Software Documentation
- [ ] Foreign Government Information

[ ] D. Distribution authorized to the Department of Defense and U.S. DoD Contractors only; (check the following):

- [ ] Foreign Government Information
- [ ] Critical Technology
- [ ] Administrative or Operational Use
- [ ] Software Documentation
- [ ] Foreign Government Information

[ ] E. Distribution authorized to DoD Components only; (check the following):

- [ ] Administrative or Operational Use
- [ ] Premature Dissemination
- [ ] Critical Technology
- [ ] Foreign Government Information
- [ ] Direct Military Support
- [ ] Software Documentation
- [ ] Specific Authority
- [ ] Proprietary Information
- [ ] Test and Evaluation
- [ ] Contractor Performance Evaluation

[ ] F. Further dissemination only as directed.

[ ] X. Distribution authorized to U.S. Government agencies and private individuals or enterprises eligible to obtain export-controlled technical data in accordance with DoD Directive 5230.25 (unclassified papers only).

CLEARED AS AMENDED
for public release

JUN 28 2011

Public Affairs
Defense Threat Reduction Agency

(minor editorial corrections)

APPROVED / JTW / 24 JUN 2011
OPSEC REVIEW/DATE

4. MATERIAL TO BE: [ ] Presented [ ] Published Date Required:

Name of Conference or Journal:

Remarks: (contd) National Security Decision-Making

CHALLENGES IN COMPUTATIONAL SOCIAL MODELING AND SIMULATION FOR NATIONAL SECURITY DECISION-MAKING

Written By:
Laura A. McNamara, Timothy G. Trucano & Charles Gieseler
Sandia National Laboratories

Contributed Essays:
Robert Albro, Mark Bevir, Jessica Glicken Turnley, François M. Hemez, Phillip Huxtable, Jeffrey C. Johnson, Lucy Resnyansky, David L. Sallach, Jean Scholtz

With thanks to:
Jennifer Perry
Project Manager
DTRA/OSRD

June 2011

This report is the product of collaboration between the Defense Threat Reduction Agency's Office of Strategic Research and Dialogues and Sandia National Laboratories, Albuquerque, NM

The views expressed herein are those of the authors and do not necessarily reflect the official policy or position of the Defense Threat Reduction Agency, the Department of Defense, or the United States Government.

Approved for public release; distribution is unlimited

Defense Threat Reduction Agency
Office of Strategic Research and Dialogues
Report Number OSRD 2011 002
Contract/IACRO Number 08-45321
Project Cost: $255,000

The mission of the Defense Threat Reduction Agency (DTRA) is to safeguard America and its allies from weapons of mass destruction (chemical, biological, radiological, nuclear, and high explosives) by providing capabilities to reduce, eliminate, and counter the threat, and mitigate its effects.

The Office of Strategic Research and Dialogues (OSRD) supports this mission by providing long-term rolling horizon perspectives to help DTRA leadership identify, plan, and persuasively communicate what is needed in the near term to achieve the longer-term goals inherent in the agency's mission. OSRD also emphasizes the identification, integration, and further development of leading strategic thinking and analysis on the most intractable problems related to combating weapons of mass destruction.

For further information on this project, or on OSRD's broader research program, please contact:

Defense Threat Reduction Agency
Office of Strategic Research and Dialogues
8725 John J. Kingman Road
Ft. Belvoir, VA 22060-6201

OSRDInfo@dtra.mil

TABLE OF CONTENTS

EXECUTIVE SUMMARY **5**

KEY WORKSHOP THEMES AND FINDINGS 6
RECOMMENDATIONS 8

1. INTRODUCTION **13**

1.1. CONTEXT: COMPUTATIONAL SOCIAL SCIENCE AND DECISION-MAKING 13
1.2. WORKSHOP BACKGROUND 14
1.3. WORKSHOP GOALS 16

2. WORKSHOP PLANNING, PARTICIPANTS, AND PROCESS **17**

2.1. WORKSHOP PARTICIPANTS AND PAPERS 19
2.1.1. Paper Contributors and Topics *19*
2.1.2. Discussants *23*

3. WORKSHOP FINDINGS **23**

3.1. CONCERNS ABOUT APPLIED COMPUTATIONAL SOCIAL SCIENCE 24
3.2. COMPUTATIONAL SOCIAL SIMULATION AS DISTINCT FORM OF COMPUTATIONAL SCIENCE 25
3.3. MODELING, SIMULATION, AND DIVERSITY IN THE SOCIAL SCIENCES 26
3.4. COMPUTATIONAL MODELS OF CULTURE 28
3.5. PREDICTION AND SIMULATION CREDIBILITY 29
3.6. MODELS, PREDICTIVE AND OTHERWISE 31
3.7. VERIFICATION AND VALIDATION 32
3.8. USABILITY AND UTILITY IN COMPUTATIONAL SOCIAL SCIENCE 33
3.9. ETHICS, MODELING, AND DECISION-MAKING 35
3.10. DATA, ERROR, AND SOCIAL NETWORKS 36

4. RECOMMENDATIONS **38**

5. CONCLUSION **42**

6. LIST OF APPENDICES **43**

APPENDIX A: ABRIDGED VERSION OF MCNAMARA-TRUCANO PAPER **45**

APPENDIX B: WORKSHOP AGENDA **75**

APPENDIX C: WORKSHOP PAPERS **81**

ROBERT ALBRO, MINING SENTIMENTS: COMPUTATIONAL MODELING AND THE INTERPRETIVE PROBLEM OF CULTURE 83
MARK BEVIR, MAKING SENSE OF COMPUTATIONAL MODELS 109
FRANCOIS M. HEMEZ, IS "PREDICTABILITY" IN COMPUTATIONAL SCIENCES A MYTH? 131
PHILLIP HUXTABLE, LEVERAGING COMPUTATIONAL SOCIAL SCIENCE FOR NATIONAL SECURITY 157
JEFFREY C. JOHNSON, GRAVE ERROR: TOWARDS A SCIENCE OF ERROR IN SOCIAL NETWORK ANALYSIS AND MODELING IN THE NATIONAL SECURITY CONTEXT 169
LUCY RESNYANSKY, SOCIAL MODELING AND SIMULATION FOR NATIONAL SECURITY DECISION-MAKING: ETHICAL CHALLENGES 189
DAVID SALLACH, SOCIAL SCIENCE MODEL VALIDATION: DOMAIN-SPECIFIC CHALLENGES 207
JEAN SCHOLTZ, A USER-CENTERED APPROACH TO SOCIAL MODELING AND SIMULATION FOR DECISION-MAKING 227
JESSICA GLICKEN TURNLEY, ASSESSING THE 'GOODNESS' OF COMPUTATIONAL SOCIAL MODELS 241

APPENDIX D: PARTICIPANT BIOGRAPHICAL STATEMENTS **255**

Executive Summary

On October 26th and 27th, 2010, Sandia National Laboratories (SNL) organized an interdisciplinary workshop in which participants from a range of institutions and research backgrounds presented and discussed papers on a range of topics related to the development and use of **computational social science** (CSS) in national security decision-making. Computational social science refers to the use of computational modeling and simulation approaches, including agent-based, social network, discrete event, and systems dynamics methodologies, to study behavioral, cultural, and social dynamics. CSS has long roots in computer science, artificial intelligence, and quantitative social science. Over the past decade, CSS methods have captured the attention of the national security community as a source of analytic and decision-support technologies for a range of challenges, from counterinsurgency to terrorism.

The workshop described in this report was part of a larger effort to identify and assess the major challenges for computational social science technologies to be effectively and responsibly incorporated into high-consequence, national security studies. Jennifer Perry of the Defense Threat Reduction Agency's Advanced Systems and Concepts Office (DTRA ASCO) sponsored Laura McNamara and Timothy Trucano of the Computing Research Center at Sandia National Laboratories (SNL) to conduct a multi-phase study to evaluate if and how computational modeling and simulation projects and technologies can add value to high-consequence, national security decision-making.

The first phase of this project was a comparative, interdisciplinary review of literature related to applied computational modeling and simulation in both the social and physical sciences. In this review, we identified three classes of challenges for computational social science in national security. We summarized these complex and interrelated challenges in three general categories: *prediction*; *verification and validation* (V&V), and *usability and utility*. These challenges are described in a summary paper that McNamara and Trucano authored (Appendix A).

The workshop described in this report was the second phase of this project. We assembled an interdisciplinary panel of experts to review our arguments in light of their own research and work experiences, and then respond, challenge, and/or elaborate upon the observations and arguments we made in the McNamara-Trucano paper. We held the workshop at the La Fonda Hotel in Santa Fe, NM (see Appendix B). It was moderated by McNamara and Trucano, with assistance Charles Gieseler, also of Sandia. Sixteen people participated in the workshop, including McNamara, Trucano, Giseler, and Perry. Ten of the participants provided draft papers to the conference; nine of the revised papers are included in Appendix C of this report.

Our participants included social, computational, and physical scientists from a range of government, industry, and academic institutions. Most, but not all, had also participated in projects to develop and deploy computational models and simulations of social phenomena for decision-making; and over half the participants had worked on computational modeling and simulation projects in in national security contexts. To help the participants frame their papers, McNamara and Trucano sent each a copy of the summary paper. We also asked each participant to address themes related to the challenges that we identified, including:

- **Prediction**: What does "prediction" mean in the social, cultural, and behavioral domains? How can we know if a model is "predictive?" Do models always entail some form of "prediction?"
- **Verification and Validation**: To what extent can V&V processes and principles from other fields, such as nuclear weapons certification and operations research, be applied to computational social science? Where do they break down?
- **Modeling and Simulation in Social vs. Physical Science:** How do computational social models and simulations differ from those in the physical sciences; and what are the implications of this difference for developing and applying models and simulations for decision-making?
- **Social, Cultural, Behavioral Data in Modeling:** What kinds of social science data are most suitable for developing and evaluating computational social models and simulations? How can social knowledge and/or information about social phenomena be translated into computationally tractable format; and what kinds of data are excluded from modeling and simulation projects?
- **Modeling, Simulation, and Decision-making:** What kinds of information do computational models and simulations generate? How should that information be communicated to consumers, and how can we responsibly and usefully balance CSS models with other forms of information? Where should models and simulations be positioned in relation to national security decision-making and, potentially, policy discussions?
- **Human Users**: What design processes and principles can ensure that modeling and simulation technologies support human information processing, and how do we evaluate the efficacy of these technologies in human cognitive processing?
- **The Ethics of Modeling and Simulation:** Given that most computational modeling and simulation efforts are multi-year, multidisciplinary projects that may or may not involve members of the user community, who is responsible for the impacts of these technologies on decision-making processes and outcomes?

Participants developed position papers in response to one or more of these themes, and we presented and discussed these papers at the workshop.

Key Workshop Themes and Findings

The workshop discussions were intense and wide-ranging, making them difficult to summarize. However, the workshop discussions revolved around the following themes:

The intersection of social science and computational modeling and simulation is producing novel and exciting interdisciplinary work. Computational modeling and simulation methods offer social scientists a portfolio of novel approaches to studying social phenomena. In addition, the challenge of modeling human society, culture and behavior is also drawing practitioners from other fields, including physics, engineering, and computer science, into the social sciences. This is an interdisciplinary confluence with tremendous implications for research practice in the social sciences, and for our ability to make sense of complex human phenomena.

Computational social modeling and simulation plays a different role in research environments than it does in decision-making applications. Researchers use computational modeling and simulation technologies to

aggregate data, examine patterns, and develop and experiment with novel explanations. Modeling and simulation can play an important role in the development of social, cultural, and behavioral theory. While these insights can be useful for government policy and decision-maker in some circumstances, these government communities are typically less concerned with theory development than with situational awareness, strategic and tactical planning, and interventions to effect desired outcomes. Moreover, models developed for application are often intended as "tools" to support analysis and decision-making in user communities that are organizationally separate from the developer community. In addition, government policy and decision-making have direct consequences for public stakeholders. These and other factors distinguish the "research" use case from the "application" use case in ways that have implications for the construction and deployment of modeling and simulation technologies.

Computational social science projects can be difficult for even subject matter experts to understand and assess. Computational social science is a new interdisciplinary field, and its applications are seemingly endless and open-ended. This makes it difficult for external observers and stakeholders to critically assess both the potential benefits and pitfalls of proposed projects and their envisioned applications.

Computational social science in national security lacks adequate input from the social sciences. Computational social modeling and simulation projects funded by government agencies are often treated as modeling projects first, as social science second. This creates tremendous risks when modelers do not understand the epistemic (lack-of-knowledge) uncertainty associated with selecting and applying social/behavioral theory, collecting, analyzing and interpreting data, and/or mitigating sources of error associated with social science problems and methods. This creates risks for decision-makers who incorporate modeling and simulation outputs in their assessments and plans.

Not all problems require modeling and simulation. Modeling and simulation is hugely popular in the national security community for a variety of social and historical reasons. When every problem is treated as a modeling problem, decision makers have less access to knowledge, methods and approaches that might provide better insights into areas of concern.

Social science problems are fundamentally different than physical problems. Social science problems concern human actors. Humans perceive, interpret, and act on the world. Human perception, interpretation, and action are necessarily contextual, historical, and specific. This limits the degree to which "general" principles of human and social behavior can be invoked to predict historically, geographically, and socially distinct processes and events in the absence of data.

However, useful and provocative analogies can be drawn to computational science in the physical and natural domains. While the target domains of computational social modeling and simulation pose tremendous epistemological and practical challenges for scientists, many of the principles involved in developing, evaluating and deploying quality computational modeling and simulation technologies are discipline-agnostic. Software engineering practices, for example, are critical to ensure that bad code does not produce erroneous outputs. Another example is lack-of-knowledge uncertainty, which is present in all science. Even physical scientists dealing with "first principles" face uncertainty in model selection and implementation. The robust physics and engineering discourse on analyzing and managing uncertainty may provide important insights for computational social modeling and simulation in national security decision-making. Comparing and contrasting computational social modeling and simulation to modeling and simulation in other domains provokes important insights into the nature of modeling and simulation.

Prediction and its many synonyms (e.g., forecasting, anticipation) are neither understood nor well defined in the social, cultural, or behavioral domains. Even practitioners in this domain do not agree on what prediction means, though many want to claim that modeling enables prediction. The fact that "predictions" are rarely well-specified makes it difficult for consumers of modeling outputs to understand what a model is "predicting" (or forecasting, or anticipating) and to hold models and their creators accountable for claims to foresight. In addition, differentiating between predictive and non-predictive uses of modeling is a difficult challenge, and one that generated a great deal of discussion among the workshop members.

Computational social models and simulations are artifacts, but they are also processes. Models and associated simulations are computational reifications of limited human knowledge about the world. Creating a model has benefits beyond the computational artifact; the process forces people to specify assumptions, identify disagreements, formalize tacit knowledge, and identify gaps in knowledge and data. The process of modeling may be more valuable than the model itself.

Who are the "users" of models and their outputs? Computational social modeling and simulation technologies often encode domain-specific forms of expertise, as well as assumptions about the entities and processes being modeled. Whether individuals without expertise in modeling, simulation, and/or social science are able to use these technologies to meaningfully and responsibly analyze complicated social processes for high-consequence decision making environments is questionable. User-oriented design, interaction design, user experience, participatory modeling, and training may help address this problem, though identifying "users" in complex government bureaucracies is often a difficult challenge. Ironically, social science research methods, such as ethnography, may be tremendously helpful in designing these technologies for human users.

Model validation in the social sciences is important and difficult. A number of factors contribute to this difficulty, from the epistemological challenges of the social sciences, to the practicalities of identifying and gathering validation quality datasets, to the time and resource requirements for conducting validation studies when models are oriented toward rapidly-changing decision spaces.

RECOMMENDATIONS

The workshop convened in Santa Fe explored a broad range of issues related to computational social modeling and simulation, from the impact of new computational techniques on social science research to the practicalities of tool adoption among intelligence and military analysts. In the wake of the workshop, McNamara and Trucano reviewed the insights, commentaries, and tremendously good work of the participants who brought their ideas and papers to Santa Fe, and distilled the following recommendations.

RECOMMENDATION ONE: Design, implement, and assess computational social science projects as hybrid, interdisciplinary research and development efforts.

We assert that computational social modeling and simulation projects should be analyzed and studied as the interdisciplinary processes and technologies that they are. Three domains in particular are germane to any assessment of computational social science models and simulations: *social science*,

computational science, and *decision support and analysis software*. When computational social science technologies are applied in decision-making, these domains intersect, as depicted in the diagram below:

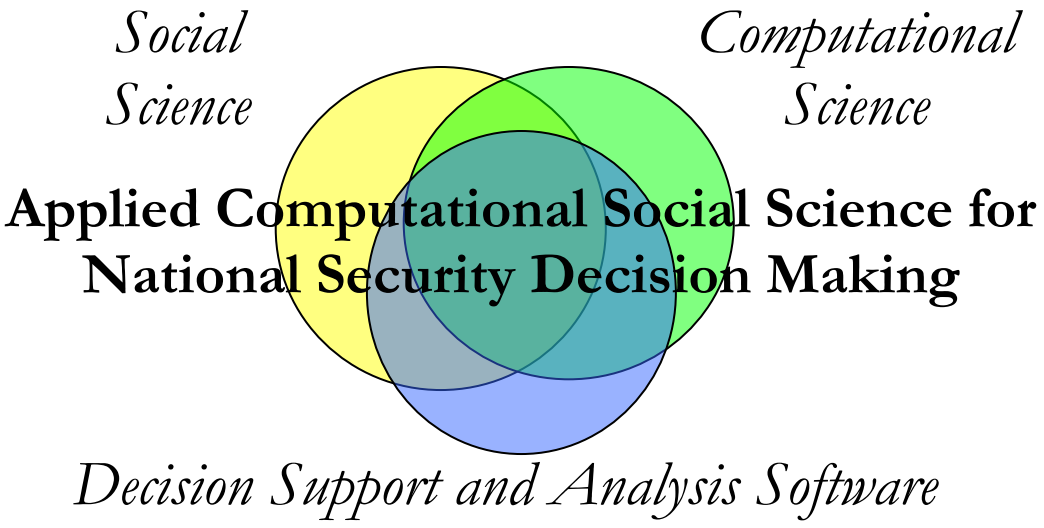


Figure 1: Applied CSS as Interdisciplinary Domain

Only an interdisciplinary framework that addresses the multiple domains of research and practice that pertain to these technologies has any hope of ensuring complete assessments of their quality and correctness. The next four recommendations derive from this intersection, and pertain to the design, development, and evaluation of computational social science projects in decision-making environments.

RECOMMENDATION TWO: Evaluate computational social modeling and simulation projects as a form of *social* science.

Computational social science is a first and foremost a form of social science. Modeling and simulation is a field of methodological research for studying social, cultural, and behavioral phenomena. Its application is only as "scientific" as the research design in which it is embedded.

Decision- and policy-makers who are attempting to make sense of a particular computational social modeling and simulation effort should ask questions about scientific validity of the theoretical and conceptual framework that underpins the model. They should ask how the modelers identified and assessed existing research on the domain, how they connect their work to pertinent questions in the social sciences, and to explain how they identified questions that the model is seeking to answer. Modelers should also be able to specify and justify the role of modeling in their research design. They should speak credibly to the challenging problems of data: how they obtained data, if these datasets are primary or secondary data, who collected the data, how, and why; to justify why data are relevant to the problem, and to discuss pertinent assumptions, limitations, and sources of error in

the data. They should be able to explain why particular modeling and simulation approaches are necessary for examining this domain, and to articulate the basis for claims to the credibility of their simulation outputs.

RECOMMENDATION THREE: Evaluate computational social models and simulations as a form of *computational* science.

As important as it is to couch computational social modeling and simulation projects in the social sciences, it does not go far enough. Social science does not ask questions about algorithms or underlying mathematical issues, nor does it point to issues of software implementation, testing, and performance. However, the field of computational science does, and framing computational social science as a form of computational science raises a second set of issues for evaluating these projects. Decision makers should ask about software engineering practices, documentation, and testing, as well as how the modeling team is evaluating the correctness and performance of the software it has written; how sources of error in the code are detected and mitigated. At a deeper level, the experience of other fields may provide valuable input to the challenges and issues that arise when computational modeling and simulation outputs are being used as a source of predictive information in decision-making, when knowledge is lacking.

RECOMMENDATION FOUR: Evaluate computational social models and simulations as *decision support tools* for individual and organizational use communities.

Tools afford efficacious human action: they are *usable and useful*. They fit well into contexts of use: they are *adoptable*. Technologies that meet these criteria are likely to be used. Definitions of usability, utility and adoptability that come from the "seller" of the model are not acceptable. Setting standards for usability, utility, and adoptability is a task that belongs squarely in the domain of the client or consumer. We assert that design processes for these technologies should involve some attention to the *intended areas of application* (this is necessary for validation *and* utility) and the *intended user communities*, if only to ensure that the resulting technologies are both usable and useful. Participatory approaches that involve users in the development process are more likely to produce educated consumers/users of both the tool *and* the information it generates. In addition, human-computer interaction, human factors, and cognitive psychology can be leveraged to develop studies that assess the impact of modeling and simulation technologies on how people assess and draw conclusions from complicated and ambiguous datasets.

RECOMMENDATION FIVE: Support interdisciplinary exchanges that enable computational social science researchers, developers, adopters, proponents, users, and stakeholders to learn how other fields analyze and evaluate models and simulations.

While we recognize that there are significant differences between the physical and the social sciences, we continue to assert that other fields' experiences with developing, deploying, interpreting and applying modeling and simulation technologies can help organizations understand how to

develop models, interpret their outputs, combine simulation outputs with other forms of information, and assess the limitations of modeling and simulation technologies in decision-making. Models are not just tools for analysis; they are artifacts that require analysis if we are to understand how they function, how we can use them responsibly, and what their limitations are.

1. Introduction

In October 2010, Sandia National Laboratories (SNL) organized an interdisciplinary technical workshop in which participants from a range of disciplines, institutions, and research backgrounds discussed computational social science for national security decision-making. The workshop was sponsored by the Defense Threat Reduction Agency's Advanced Systems and Concepts Office (DTRA ASCO) through Jennifer Perry, and was organized and moderated by Laura McNamara and Timothy Trucano of the Computing Research Center at Sandia National Laboratories. In addition, Charles Gieseler of the Sandia National Laboratories' Cognitive Modeling Department participated in the discussions and assisted with workshop logistics and recording exchanges among participants. This workshop was part of a larger project entitled "Grand Challenges in Modeling and Simulation for Decision Support" that McNamara and Trucano had been pursuing in conjunction with Perry since October 2009.

This report presents the workshop's findings. In the following pages, we provide some context for the workshop, describe our approach to organizing and executing the event, and the major themes and conclusions that emerged from our contributors' papers and discussions. Our participants' papers and biographical statements are included as appendices to this report.

1.1. Context: Computational Social Science and Decision-Making

Since the early 1990s, computational social science (CSS) has emerged as a major field of interdisciplinary research, with practitioners from the social, computational, mathematical, and physical sciences exploring a range of approaches to studying social phenomena through modeling and simulation.[1] CSS refers to the use of computational modeling and simulation approaches, including agent-based, social network, discrete event, and systems dynamics methodologies, to study behavioral, cultural, and social dynamics. CSS has long roots in computer science, artificial intelligence, and quantitative social science.

In the wake of the 9/11 attacks, and the subsequent invasions of Iraq and Afghanistan, a number of institutions in the United States' national security community, and particularly the U.S. Department of Defense (DoD), have invested in CSS research. Such investments are part of a larger trend in the national security community to seek relevant information and knowledge from the social sciences to develop novel strategies for understanding and addressing complex national security challenges, particularly radicalization, terrorism, and insurgency. Among many policy and decision-makers,[2]

---

[1] By "modeling," we mean the development of an abstraction that captures elements and interrelationships in a posited system. By "simulation," we mean an active instantiation of a model that temporally and logically evolves the constituent elements, resulting in a set of outputs for analysis (see P. Bratley, B. L. Fox, and L. Schrage, *A Guide to Simulation*. New York: Springer, 1987, page 1). Most scientific disciplines, including the social sciences, rely on both modeling and simulation to develop hypotheses, formalize emerging understandings, identify and address gaps in knowledge, and to specify and refine theories. Neither modeling nor simulation requires a computer: models can be as simple as a mathematical equation, while experiments are a form of simulation.

[2] We differentiate between policy and decision-makers at the request of our DTRA sponsor. In this paper, policy-makers denote those elected and appointed officials who set broader agendas and/or goals for federal government departments and agencies. Decision-makers are generally the employees within those departments and agencies who are charged with implementing policy. Many of the computational modeling and simulation projects and technologies that we are describing in this paper are oriented toward decision-makers – for example, analysts and military personnel – in the Department of Defense and the Intelligence Community. However, there is no reason in principle that computational social models and simulations might not be used in policy analysis and development.

computational social models and simulations are perceived as having a wide range of potential applications, from training military personnel to tactical and operational support for in-field decision-making.

At the same time, CSS is a relatively new field of interdisciplinary research and development. In particular, it represents the intersection of multiple deep areas of research, driven primarily by the increasing power and availability of computational technologies as well as new sources of data (internet and wireless communications data, for example). Computational modeling and simulation methodologies are fields of study in their own right, as are the heterogeneous disciplines of the social sciences. Moreover, CSS projects draw on knowledge and methods from other fields of study, including graph theory, information visualization, and statistical physics. Because of the field's heterogeneity and newness, people without experience in computational modeling and simulation – even those with expertise in the target phenomena being modeled (e.g., geographic and cultural regions, religious movements) – can find it extremely difficult to assess the quality of a particular computational modeling and simulation effort. In short, the field is new, its disciplines diverse, and its applications seemingly endless and open-ended, making it difficult for external observers and stakeholders to critically assess both the potential benefits and pitfalls of proposed projects and their envisioned applications.

1.2. Workshop Background

This problem – the absence of a framework for assessing the state of computational social modeling and simulation as an applied science, and for identifying and posing evaluative questions about computational modeling and simulation projects for government agencies – was precisely the challenge that the DTRA/Sandia project sought to address. In the Fall of 2009, we began a research project U.S. to address the following questions:

- What are the technical challenges that need to be addressed before computational social models are ready for "prime time" application in national security decision-making environments?
- How might these challenges be addressed? To what extent can we apply "lessons learned" from other scientific communities? How do social science theory and methods address these challenges? What are the ramifications of not addressing them?
- How can we promote a mutual understanding of these challenges and the issues surrounding the practical application of models in national security decision-making among all relevant communities?

From the Fall of 2009 until the Summer of 2010, McNamara and Trucano reviewed literature related to modeling and simulation, social science, and decision-making. Because the intersection of computational social science and national security is so interdisciplinary, we purposely cast a broad net when gathering our literature. Our review included selections in CSS, both within and outside the realm of national security decision-making. However, we also examined literature from fields in which computational modeling and simulation is a more established methodology, both for research purposes and organizational decision-making. For example, we examined literature on weather and economic forecasting, as well as operations research and nuclear weapons certification.

In this review, we discovered a great deal of methodological guidance on how to design, develop, evaluate, and deploy computational modeling and simulation technologies as decision-support tools. We also realized, somewhat surprisingly, that the existing CSS literature rarely calls on the practical experience of other fields, such as weather forecasting and nuclear weapons certification, whose practitioners have developed rich understandings about the incorporation, evaluation, benefits and limitations of modeling and simulation technologies in forecasting and decision-making activities.

We concluded that computational social modeling and simulation demands significant methodological development if its technologies are to be used as tools to support high-consequence decision-making. Inherent in this statement are three arguments: first, computational modeling and simulation technologies always entail limited approximations to the real world. However, as computational models and simulations become more sophisticated, limitations can be difficult to identify and evaluate. Second, the demands and requirements of decision-making environments, such as those found in the U.S. Department of Defense, are different from the research environments where CSS is primarily rooted. Decision makers in the national security community have problems to understand, plans to make, and resources to allocate; and in all of these, they may literally be dealing with life-and-death challenges. Finally, no model or simulation is inherently a "decision support tool;" instead, tool development is a distinct field of research and development that explicitly takes account of human users and organizational contexts of deployment.

From these assertions, we identified a set of challenges for computational social science as it moves from research into application in national security environments. We began with a normative assertion: all modeling and simulation projects should be approached with skepticism. However, we emphasized that skepticism must be *productive* rather than cynical or dismissive, and spoke of a pragmatic, applied, constructive skepticism that would support understandings about what modeling and simulation technologies are good for; their appropriate role in helping people solve problems and make sense of complex situations; and the appropriate use of computational models and simulations in framing decision spaces and as inputs to setting courses of action.

We argued that *productive skepticism* entails three separate classes of challenges for computational social modeling and simulation. Firstly, productive skepticism does not assume that models and simulations produce analyses, provide forecasts, or make decisions. In particular, we pointed out that the *promise of prediction* is a dominant theme in the discourse about the benefits of computation social models and simulations. However, we drew on the work of a number of science policy experts to make the argument that scientific prediction and policy prediction are substantially different enterprises; and moreover, the temporal, geographical, and social dimensions of sociobehavioral prediction for policy making are rarely well specified. When applied to "prediction," productive skepticism requires careful specification of a prediction's content, boundaries, and limitations.

Secondly, productive skepticism emphasizes the importance of rigorous evaluation, along the lines of what has been done in *verification and validation* (V&V) in fields such as computational physics and operations research. In these fields, V&V encompasses a set of methods to *evaluate how well a model/simulation meets the demands of an intended application*, not to determine whether or not a simulation is "correct." V&V methodologies enable stakeholders to assess how well a particular model captures the critical aspects of the domain being modeled (validation); and how well the software is actually executing that model (verification). Verification and validation are complex challenges for any field, insofar as V&V is a rigorous evaluation process that comprises problems of conceptual/theoretical

validity, data quality, and the identification of referents or comparison points to "tie" models and simulations to the real world. However, V&V is particularly challenging for the social sciences, because of the heterogeneity of theoretical frameworks and the interpretive, representative nature of social science data. Moreover, V&V is also organizationally and financially challenging, because V&V activities can require significant investments in time, money, and personnel – not easy to come by in rapidly shifting, demanding decision-making environments such as those found in the U.S. Department of Defense.

Thirdly, to say that a model supports decisions or is "predictive" elides the presence of the human being(s) who are developing, running, and interpreting modeling and simulation outputs and ultimately making predictions. Accordingly, we argued that productive skepticism does not take human users for granted, but is focused on issues of design and evaluation to understand how modeling and simulations methodologies can support human information processing and organizational decision-making structures. Ensuring *usability and utility* requires attention to issues of human factors, ergonomics, and distributed human cognitive systems, such as those found in high-consequence decision-making environments. Whether or not a "tool" supports humans in these environments is a complicated design and evaluation problem above and beyond the internal and external correctness of a computational modeling and simulation technology.

Lastly, in addition to these three primary challenges, we noted that the application of CSS technologies in high-consequence decision-making raises ethical challenges as well. Just as computational social science represents the intersection of a number of theoretical and methodological research areas, the ethical challenges of computational social science are likely interdisciplinary as well. We believe that the ethical frameworks developed in computer science may provide some guidance for applied CSS, but that ethical principles in the social sciences are likely germane to the development of the field as well. Researchers in a number of social science fields, including anthropology and psychology, have concluded that certain forms of applied social science for national security problems are ethically problematic. However, these discipline's analyses typically do not address the specifics of computational social models and simulations as a form of applied social science.

In the end, we gathered over 900 books and articles and drafted a 70-page draft review, addressing this diverse set of themes. This review was submitted to DTRA ASCO in July of 2010 (an abridged version of this longer paper is attached as Appendix A of this report). As extensive as this review was, we recognized that it was neither sufficient nor complete. As reviewers, we saw significant connections among these different bodies of literature, but we realized that our comparisons were likely incomplete, perhaps controversial, and that we had probably missed important themes. We also wanted to collect input from experts in other fields about the challenges that accompany computational social science as it moves from research into an application domain; and hoped that we could leverage this input to develop a framework for understanding and critically assessing computational modeling and simulation projects.

1.3. Workshop Goals

The Santa Fe workshop provided an opportunity for us to discuss our primary challenges – *prediction*, *verification and validation*, and *usability and utility*, along with the ethical considerations of modeling and simulation in decision-making – with representatives of the different research areas that we had

identified in our literature review. The primary goal of the workshop was to have an interdisciplinary panel of experts review our arguments in light of their own research and work experiences, and then respond, challenge, and/or elaborate upon the observations and arguments we made in our paper. Because our literature review was so interdisciplinary, we selected experts from a range of fields who collectively could speak to intersections that we identified in our review. During the workshop, the participants examined and discussed the state-of-the-art in computational social modeling and simulation in both research and decision-making contexts. They spoke to the challenges we identified in our paper, sometimes critically; and identified other issues that face CSS projects being transitioned into real-world decision-making in national security. Most importantly, we discussed what advice we might give to decision-makers in the national security community about how to evaluate computational social modeling and simulation projects, so that stakeholders in these projects are empowered to make informed decisions about how to fund, pursue, evaluate, and apply CSS-based tools in their domains of responsibility.

2. Workshop Planning, Participants, and Process

The workshop was organized around individual papers contributed by invited participants, focusing on specific themes and topics related to computational social science in national security decision-making. In the Spring of 2010, McNamara and Trucano identified a set of focal topics to be explored by the participants. These focal topics included the current state of CSS and its applications in national security contexts, with an emphasis on *prediction, verification and validation,* and *usability and utility* when deploying CSS technologies as decision support "tools" in national security organizations.

Once we had identified these focal topics, we developed set of cross-cutting sub-topics for elaboration in specific papers. These cross-cutting themes are listed below. None are mutually exclusive, and each speaks to all three of the topics that we identified above.

- **Prediction**: What does "prediction" mean in the social, cultural, and behavioral domains? How can we know if a model is "predictive?" Do models always entail some form of "prediction?"
- **Verification and Validation**: To what extent can V&V processes and principles from other fields, such as nuclear weapons certification and operations research, be applied to CSS? Where do they break down?
- **Modeling and Simulation in Social vs. Physical Science:** How do computational social models and simulations differ from those in the physical sciences; and what are the implications of this difference for developing and applying models and simulations for decision-making?
- **Social, Cultural, Behavioral Data in Modeling:** What kinds of social science data are most suitable for developing and evaluating computational social models and simulations? How can social knowledge and/or information about social phenomena be translated into computationally tractable format; and what kinds of data are excluded from modeling and simulation projects?
- **Modeling, Simulation, and Decision-making:** What kinds of information are generated by computational models and simulations? How should that information be communicated to consumers, and how can we responsibly and usefully balance computational modeling and

simulation findings with other forms of information? Where should models and simulations be positioned in relation to national security decision-making and, potentially, policy discussions?

- **Human Users**: What design processes and principles can ensure that modeling and simulation technologies support human information processing, and how do we evaluate the efficacy of these technologies in human cognitive processing?
- **The Ethics of Modeling and Simulation:** Given that most computational modeling and simulation efforts are multi-year, multidisciplinary projects that may or may not involve members of the user community, who is responsible for the impacts of these technologies on decision-making processes and outcomes?

We used this literature review to identify researchers that we believed could bring a unique perspective on these topics. In addition, Perry commented on the thematic areas, and suggested names of possible participants that she had either worked with or encountered as a project manager at DTRA ASCO. Each individual was chosen to address one or more of the challenges we identified in our paper, drawing on their own experience to respond to and elaborate upon a set of themes. In addition, we identified discussants with expertise in the intersection of science and national security policy to provide additional context to the workshop papers and exchanges.

Our initial expert list included social scientists with experience in computational social modeling and simulation projects, policy and decision-making; computational scientists (including one physicist and a PhD engineer) with experience in developing, evaluating, and applying computational models and simulations to national security technologies, including nuclear weapons; representatives of the national security user community for whom CSS technologies are being developed; and researchers with expertise in designing, deploying, and evaluating software tools for human users in the national security community. Most, but not all, had also participated in projects to develop and deploy computational models and simulations of social phenomena for decision-making, some inside the national security community and others in fields such as development and natural resources management. In the end, we had sixteen participants: McNamara and Trucano as workshop chairs, Perry as observer, ten paper contributors, and three discussants.

McNamara began contacting potential participants in June of 2010, explaining that we had selected them because of their ability to focus on a particular problem or theme (the themes that we asked them to consider are discussed in the next section of this report). Roughly a month later, we sent each of them a copy of the McNamara and Trucano paper, and requested that they read the paper in light of the particular topic or theme that we had asked them to consider. We gave our participants leeway in their writing assignments: they could respond to our commentary, elaborate on points that we had identified, and/or bring forward ideas that we had not addressed.

The October 2010 workshop provided these experts an opportunity to review and discuss each others' papers. However, the workshop was also a forum in which these experts – many of whom were largely unfamiliar with each others' research fields – could exchange ideas about the role, development, and application of computational methodologies for advancing knowledge; as well as the translation of research technologies into applied tools. The biggest challenge of the workshop was ensuring that researchers from such diverse backgrounds would be able to speak to each other's papers and ideas, in the context of CSS for national security decision-making. To facilitate the participants' familiarity with each other's work, we asked paper contributors to submit an abstract

and a biographical statement several months in advance of the workshop. We compiled these into a single document and sent it to the entire seminar participant list, so that people would have a sense of who was attending and what the tentative topics would be. We also asked participants to submit their essays two weeks prior to the seminar, so that we could distribute them to the rest of the group for pre-reading material. To further facilitate discussion, we also reviewed each of the papers and assigned each workshop participant the role of primary discussant for another seminar participant's paper. During the actual workshop, each participant briefly presented his or her own paper, then the assigned discussant reviewed the paper, identifying key themes and raising discussion points. The exchange then opened to the entire group. The agenda is included in Appendix B of this report.

This structured seminar format worked extremely well. During the two days of discussions, the group engaged in lively and collegial discussions about each other's papers, and the exchanges ventured into a range of topics related to computational modeling and simulation, social science, policy and decision-making, ethics, and national security challenges. We are grateful to the participants for their thoughtful papers, their sincere engagement with the topics at hand, and their enthusiasm for exchanging ideas and learning from each other.

2.1. Workshop Participants and Papers[3]

In this section, we introduce our workshop paper authors and discussants. We identify the theme we asked the paper authors to consider and provide a brief summary of their papers, which can be found in full in Appendix C of this report. We also provide brief biographical statements for these experts. Full biographical statements for all of the workshop participants can be found in Appendix D of this report.

2.1.1. Paper Contributors and Topics

Jeffrey C. Johnson was asked to discuss the topic of *social science data collection, management, and use in computational social science projects for national security decision-making.* Johnson is a quantitative anthropologist and is Senior Scientist at the Institute for Coastal Science and Policy, and University Distinguished Research Professor in the Department of Sociology with adjunct appointments in Biology, Anthropology, and Biostatistics at East Carolina University. He is also the Social Science Program Manager for the U.S. Army Research Office where he is developing a basic scientific research program in the social sciences. Johnson's workshop paper is a warning against taking social science "data" for granted. He acknowledges the importance of good theory selection in developing computational models of social dynamics, but emphasizes that the collection and analysis of social science data presents significant methodological challenges as well. In particular, Johnson argues that greater focus on error in social network modeling and analysis is necessary if computational analysis is to be incorporated into real-world decision-making. Johnson identifies roughly a dozen types of error that can threaten data and model validity in the context of social network analysis and

[3] In addition to those experts listed below, we had also invited **Simone Youngblood,** currently at the Johns Hopkins Applied Physics Laboratory and formerly of the U.S. Defense Modeling and Simulation Office (DMSO) to discuss the history and importance of Verification and Validation, and to reflect on the fiscal, political, and organizational challenges of introducing rigorous V&V into modeling and simulation projects. Youngblood initially accepted the invitation, but later withdrew from the workshop, citing family obligations and scheduling pressures.

assesses how different errors can lead to significantly different analytical conclusions. Johnson calls for a science of error so that stakeholders in network models can better understand the impact of error on model performance and real-world decisions.

David L. Sallach was asked to examine the topic of v*erification and validation to identify challenges that are specific to the domain of social science*. Sallach is a sociologist who specializes in applying agent modeling and simulation technologies within social science domains, drawing on social theory to develop trans-scale social models. He received his PhD from the University of Nebraska at Lincoln; has taught both sociology and computer science, and served as Director of Social Science Research Computing at the University of Chicago from 1998-2003, where, along in conjunction with Nicholson Collier, he designed the architecture of the Repast agent simulation toolkit. Since 2003, he has been Associate Director of the Center for Complex Adaptive Agent Systems Simulation at the Argonne National Laboratory, and also a Senior Fellow at the Computation Institute at the University of Chicago and Argonne National Laboratory. In his paper, Sallach critiqued the implication that verification and validation frameworks from other fields provide a sufficient guide for formal validation efforts in CSS projects. Instead, he argues that social science problems present unique challenges for model validation above and beyond those in the physical sciences, such as context effects, endogenous and emergent behaviors in social systems, and the hidden nature of intent, all of which are present in social dynamics, and which are typically absent from V&V research in fields like physical science and operations research.

Jessica Glicken Turnley has experience developing CSS projects in interdisciplinary settings, and was asked to discuss the problem of *evaluating the quality* (what she refers to as "goodness") *of a modeling and simulation project*. She is President of Galisteo Consulting Group, Inc., a consulting firm in Albuquerque, NM. She is formerly a member of the Defense Intelligence Agency Advisory Board and currently holds an appointment as Senior Fellow, Joint Special Operations University, USSOCOM, where she provides research, analysis and concept development of selected special operations issues, computational modeling and simulation initiatives, and organizational and cultural topics pertinent to both U.S. national security organizations and adversaries. Turnley's paper draws on her experience observing and participating in national security-related modeling and simulation projects to examine how organizations use these in decision-making. The complexity of sociocultural phenomena, the contingency of data, the timescales of social phenomena, all mean that computational social models and simulations are unlikely to be "predictive" in the way that models/simulations of physical phenomena are assumed to be. Turnley believes this has implications for what "verification" and "validation" entail in this domain, and argues that the V&V literature in the physical sciences emphasizes predictive V&V. She suggests that evaluation methodologies used for CSS technologies should be expanded beyond traditional prediction-oriented V&V to address a broader range of intended applications.

Lucy Resnyansky was tasked to examine the *ethical challenges of developing, deploying, and using computational social modeling and simulation technologies in high-consequence decision-making*. Resnyansky is Research Scientist with the Defence Science and Technology Organisation (DSTO) Australia. She has a Bachelor (Honors) degree in Linguistics (1985) and a PhD in Social Philosophy (1994) from Novosibirsk State University (Russia); and a PhD in Education (2005) from the University of South Australia. Resnyansky conducts research in areas of social modeling, national security and intelligence analysis, interdisciplinary research; sociocultural implications of technology; Internet-mediated social interaction; activity theory; and social semiotics. In her paper, Resnyansky argues that the rapid rise of information and communications technologies has raised significant ethical

challenges for national security decision makers, revolving around freedom and surveillance. The extent to which the ethical challenges of communications and information technologies are present in the domain of computational modeling and simulation may depend on how information and communications data sources are incorporated into models. However, she also points out that computational modeling and simulation technologies present additional ethical dilemmas related to the role of models in translating expert knowledge for non-expert users, and the quality and outcomes of organizational decision-making processes. She argues that conscientious research, design, and development activities in modeling and simulation are critical to address the ethical challenges of applying simulation technologies in real-world decision spaces.

Francois M. Hemez was asked to examine *the limitations of using simulations as predictive tools under conditions of uncertainty*. Hemez has been Technical Staff Member at Los Alamos National Laboratory since 1997. He was a member of the Weapon Response group (ESA-WR) for seven years; served as ESA-WR Validation Methods team leader for one year; and is currently with XTD-Division. He has managed the verification project of the Advanced Scientific Computing program for two years and currently manages the Predictive Capability Assessment project, while contributing to the development and application of verification and validation (V&V), uncertainty quantification and decision-making for engineering, nuclear energy and weapon physics projects. Hemez' paper was unique among the contributions, because Hemez leveraged his work in physical modeling and simulation to address issues in the domain of social modeling and simulation. Hemez points out that uncertainty in modeling and simulation can derive from the inherent randomness of a phenomenon; or it can stem from the assumptions we make in constructing a model – assumptions that may be necessary to render a phenomenon model-able, but that also hide our lack of knowledge and understanding. This complicates "prediction," leading Hemez to argue that modelers face tradeoffs among fidelity-to-data, robustness to lack of knowledge, and consistency of numerical predictions. For Hemez, "predictability" is a broad question that addresses the consistency of predictions across classes of models that are demonstrably equivalent in their a) accuracy and b) robustness to lack-of-knowledge. In making this argument, Hemez seeks to widen the discussion about model "correctness" to build a more rigorous mathematical and practical foundation for using models, while understanding their limitations for forecasting under conditions of uncertainty.

Jean Scholtz has worked in the area of user-centered evaluation of technology for 20 years, and was tasked with examining the *challenge of human factors design, deployment, and evaluation in computational modeling and simulation projects*. She is Chief Scientist at Pacific Northwest National Laboratory, working part time in user-centered evaluation of visual analytic systems, with expertise in law enforcement and intelligence analysis contexts. Previously, she worked for the National Institute of Standards and Technology where she developed metrics and methodologies for user-centered evaluations of programs for the U.S. Defense Department and the Intelligence Community. In her paper, Scholtz argues that user-centered approaches are lacking in the design and deployment of computational social modeling and simulation technologies. Drawing on her extensive research in user-oriented metrics for usability and utility of intelligence analysis software, Scholtz emphasizes that the translation of prototype technologies into usable and useful tools must begin with studies of the user communities. Effective end-user metrics not only assess the utility of the technology and its potential adoptability, but its impact on the modeling and simulation results, and perhaps the analysis products, that users create and pass to decision-makers.

Mark Bevir is a Professor in the Department of Political Science at the University of California at Berkeley. His primary research interests are in political theory (including the history of political

thought, political philosophy, and the philosophy of the human sciences) and public policy (including interpretive analysis, organizational theory, and governance). His research examines the intersection of social science, especially political theory, with real-world governance, a topic on which he has published over a dozen books and 150 articles. For this workshop, Bevir's paper addressed the interpretive and contextual nature of social science vis-à-vis the physical sciences, drawing a distinction between "naturalism" and "anti-naturalism" in the physical and social sciences. For Bevir, this distinction has significant implications for how we should approach individual modeling projects, as well as the broader program of computational social modeling. Specifically, Bevir points out that *theory in the social sciences tends to be more socio-historically contextual than theory in the physical sciences; and that the contextual, specific nature of both social phenomena and the explanations we develop for them make "prediction" an unwieldy goal for computational social modeling.* Bevir then discussed the implications for computational modeling and simulation, forecasting, and governance, emphasizing the importance of narrative in helping decision makers locate actors and their actions in broader historical, geographical, and political contexts.

Robert Albro is a sociocultural anthropologist and Professor of International Communication at the School of International Service at American University. He has been conducting research on popular and indigenous politics in Bolivia since 1991 and is currently studying how "culture" is being used in national security. Dr. Albro's research and writing have been supported by the National Science Foundation, Mellon Foundation, Rockefeller Foundation, and the American Council for Learned Societies, among others. Dr. Albro has also been a Fulbright scholar, and has held fellowships at the Carnegie Council for Ethics in International Affairs, the Kluge Center of the Library of Congress, and the Smithsonian Institution. Albro recently served on National Research Council's Committee on Unifying Social Frameworks and teaches in the School of International Service at American University. Albro used his paper to *examine the concept of "culture" as involved in the social sciences, and as deployed in the context of computational social modeling and simulation in national security*. In doing so, Albro identified significant disconnections between the way that cultural anthropologists and computational social modelers and simulation proponents approach the study of culture. Not only does this raise questions about how to assess the validity of cultural data, but highlights how computational technologies may exclude forms of information and knowledge, such as narrative, that may be critical for decision-making but unwieldy in computational projects.

Phillip Huxtable received his PhD in Political Science in 1997 from the University of Kansas, with a focus on African Politics and quantitative analysis. Since joining the U.S. Department of Defense in 1999, he has filled a variety of roles, leading teams focused on incorporating the theories and methods of the social, engineering, and physical sciences into analysis. Huxtable currently leads a team of senior scientists and engineers charged with assessing emerging analytic technologies for their potential relevance to future national security issues. In his paper, Huxtable describes rising interest in computational social science among national security and defense communities. However, he points out that *academic social science, including its computational form, is not well tuned to the demands of government decision-making.* He suggests that government agencies examine how they invest in, plan, and evaluate social science projects, emphasizing that multi-disciplinary, collaborative approaches can facilitate the adoption of new capabilities into existing analytic communities.

Michael Vlahos is Professor of Strategy at the United States Naval War College. He is the author of *Fighting Identity: Sacred War and World Change*, an analysis of how war — as sacred ritual — shapes collective identity. His career includes service in the Navy, the CIA, Johns Hopkins, and the State Department. Dr. Vlahos submitted a brief paper and provided insightful comments during the

seminar, but withdrew his paper from publication. His remarks are noted in the transcript of the workshop and we have included some of his comments in the summary report.

2.1.2. Discussants

In addition to the paper contributors, we invited several discussants to help us review and frame the issues. Each was selected for experience in computational modeling and simulation, national security decision-making, the intersection of science and policy, or all three. Our discussants' biographical statements are below:

Alyson Wilson was invited as a discussant because of her extensive experience in developing and applying statistical models for national security problems. She is an Associate Professor in the Department of Statistics at Iowa State University and a Scientist 5 in the Statistical Sciences Group at Los Alamos National Laboratory. Dr. Wilson is the founder and past-chair of the American Statistical Association's Section on Statistics in Defense and National Security. In addition to numerous publications, Dr. Wilson recently co-authored a book, *Bayesian Reliability*, and has co-edited two other books, *Statistical Methods in Counterterrorism: Game Theory, Modeling, Syndromic Surveillance, and Biometric Authentication* and *Modern Statistical and Mathematical Methods in Reliability*. Dr. Wilson received her PhD in Statistics from Duke University, her M.S. in Statistics from Carnegie-Mellon University, and her B.A. in Mathematical Sciences from Rice University.

Gerald Epstein received S.B. degrees in physics and in electrical engineering from the Massachusetts Institute of Technology and a PhD in physics from the University of California at Berkeley. He has extensive experience working in the intersection of science, policy-making, and national security and provided a valuable perspective on the contexts in which CSS modeling and simulation tools are being sought and deployed. Epstein joined the American Association for the Advancement of Science (AAAS) Center for Science, Technology, and Security Policy (CSTSP) as Director in October 2009. Prior to joining CSTSP, he was Senior Fellow for Science and Security in the Homeland Security Program at the Center for Strategic and International Studies, where he worked on reducing and countering biological weapons threats and improving relations between the scientific research and national security communities. Dr. Epstein is helping McNamara assemble a policy seminar on this topic at AAAS in 2011.

Charles Gieseler is a software engineer who is interested in human computer interaction and computational modeling. He completed a Masters in Computer Science from Iowa State University in 2005 with a focus in machine learning for agent-based computational economics. He is currently a software engineer working with the Cognitive Modeling Department at Sandia National Laboratories, where he is developing user interaction and simulation technologies for cognitive modeling applications.

3. Workshop Findings

The workshop discussions covered a wide range of topics, from the philosophical underpinnings of modeling and simulation in the natural and social sciences, to the practicalities of introducing new technologies into existing work environments. We take the energy and range of these discussions as a sign that the workshop was successful in provoking fruitful exchanges, but it does make

summarizing the key themes a bit challenging. The summary below pulls out some themes that occurred repeatedly in the discussions.

3.1. Concerns about Applied Computational Social Science

The workshop participants expressed both excitement and concern about the explosion of research and interest in computational social science. Several of the participants with expertise in computational social modeling and simulation described important advances; e.g., in social network analysis and modeling, while others argued that computational methodologies have the potential to dramatically advance the social sciences. All agreed that generous national security investments are playing an important role in the development of CSS methodologies; however, the application of these technologies in real-world environments was a source of concern for all the participants.

Since the 9/11 attacks, there has been an increase in the number of U.S. government-funded CSS projects oriented toward topics of interest to national security policy and decision makers. Indeed, one of the major themes of the workshop was the impact of national security funding on the field. The second day of the seminar opened with a long discussion about the relationships among academic social science, applied and computational social science, and national security decision-making. Workshop participants agreed that national security and defense investments through agencies like the Defense Advanced Research Projects Agency (DARPA) among others can have a huge and positive impact on a field's growth. However, such large investments can quickly run ahead of the state-of-the-art in academia, making it difficult for professionals outside of the defense trajectory to keep track of how computational social science is evolving in these domains. Moreover, as several of the seminar participants pointed out, computer scientists have an advantage over their social science colleagues because they have a stronger tradition of interacting with the national security community.

As a result, the "computational" part of the broader "computational social science" community may have a stronger presence in national security discussions than do formally trained social scientists. Indeed, many of the participants commented on the absence of social science expertise, both qualitative and quantitative, in computational social modeling and simulation projects. One participant related an experience at a CSS conference: when she asked how many people in the room had training in the social sciences, she discovered that almost all the workshop participants were computer scientists. As she pointed out, it is difficult to evaluate the quality of a model and/or simulation without some basic understanding of the target domain.

Along these lines, several workshop participants expressed the opinion that computational social modeling and simulation technologies are being oversold as predictive technologies. One of the participants observed that military and intelligence clients are so interested in predictive capabilities that they tend to invest in projects that claim predictive goals, without necessarily considering what prediction means in the context of social phenomena. Another participant observed that these clients rarely specify what they expect from a model. She pointed out that expectations for a 'good' model are not clearly spelled out they are unlikely to be met, which jeopardizes future investments in the field. Both these participants agreed that educating the client community about what computational social models and simulations can and cannot do is an important challenge for the CSS community.

The issue of transitioning tools to users was a recurrent theme as well. Not only do many CSS practitioners lack formal training in the social sciences, but social scientists rarely have training in software development. This raises problems when computational techniques are being advertised to defense and intelligence analysts as "tools." At a very basis level, scientifically "correct" modeling and simulation technologies may not meet user requirements for usability or utility. This is not unique to computational social science. Software developers often do not appreciate how new technologies impact people's work practices in ways that may be awkward or troublesome from the user's perspective. When new technologies flop, it is often because the proponents of new tools failed to account for important factors in the use environment, such as ensuring the traceability of users' analytic judgments to reliable data sources.

More importantly, computational modeling and simulation technologies may require significant methodological and domain experience to use them properly. Two of the participants with extensive national security community expertise observed that the interdisciplinary nature of computational social science complicates the challenge of technology transition, because CSS technologies embody deep forms of expertise that may not be easy to throw over the proverbial fence. Users may be reluctant to trust black box models that they do not understand; or, more dangerously, they may put too much faith into computational modeling and simulation technologies whose limitations they do not fully appreciate. This provoked a discussion of social network analysis, which - as Jeffrey Johnson describes in his paper - is prone to a number of errors that can threaten analytical validity. Lacking training in theory, methods, and techniques, it is easy to make mistakes that can have existential implications when network analysis products are used to support very high consequence decisions, such as targeting in operational contexts.

3.2. Computational Social Simulation as Distinct Form of Computational Science

McNamara and Trucano's provocation paper argued that applied CSS projects (such as those being pursued for national security decision-making) can leverage techniques and approaches from fields like computational physics to develop more robust approaches for evaluation and deployment. However, several of our participants pointed out that modeling in the social sciences is a different enterprise than in the physical sciences, and asserted that this has implications for issues from application to validation.

For example, in his paper, David Sallach elaborated on the McNamara-Trucano paper by engaging the domain specific aspects of the social sciences, including endogeneity, nonlinearity, sequence effects, the intentions of actors and the importance of context, that make validation very challenging in the social sciences. In doing so, he asserted that computational social modeling and simulation projects are qualitatively different from computational physical modeling and simulation projects. Social science is inherently messy, he argued, and techniques for performing and evaluating simulations in the physical sciences are unlikely to provide clever or creative ways to avoid all that messiness.

Mark Bevir's paper made similar points, but his provoked intense exchanges on the very nature of social science vis-à-vis computational modeling and the physical sciences. During the discussions, Bevir noted that much of the discussion about computational social science in the McNamara-Trucano provocation paper emphasized how models are being used. He argued that it is equally

important to understand how social modeling and simulation projects *differ* from computational models of the natural world. Bevir drew a distinction between natural (physical) and social phenomena, pointing out that modeling in the natural sciences – for example, physics, chemistry, biology, and geology – focuses on processes and objects that do not have intentionality. This has implications for modeling and simulation of social phenomena: it means that the context of human action is critical for understanding social processes, because people are acting on their interpretations of local events. Secondly, Bevir argued, social scientists are human beings engaged in the act of interpretation; and moreover, they are interpreting the actions and intentions of other human beings, as they develop descriptions, explanations, and theories. By extension, the social sciences are themselves interpretive, which means that the theoretical formulations of social science are less durable and robust than the theories that physical scientists rely upon. Other participants added that theoretical discussions in the social sciences are likely to be more contentious and problematic than in the physical sciences; to external observers, the social sciences can appear to be a disconnected jumble.

3.3. Modeling, Simulation, and Diversity in the Social Sciences

In highlighting the difference between natural and social phenomena, Bevir's paper opened a series of exchanges about the nature of knowledge in the social sciences, and the implications for how social science research, techniques, theories and data are incorporated into national security decision-making. The discussions that took place in the workshop mirrored longstanding debates among social scientists about the philosophical underpinnings of social science knowledge. Without diving too deeply into these topics – which can quickly become esoteric and confusing – we want to call attention to these debates, because they are germane to what people believe can be done with modeling and simulation in social domains. They influence how we approach the themes raised in the McNamara-Trucano paper, such as verification and validation, usability and usefulness, and predictive applications. They are also important in considering whether computational social modeling and simulation projects represent an optimal means of incorporating social science knowledge into decision-making.

Social scientists have many different views about how and if social science can establish reliable, replicable knowledge about the world, and the extent to which social realities can be studied objectively in the same way that, say, particle physicists study the elementary constituents of matter.[4] For some social scientists, the goal of social sciences is to account for the real world as objectively as possible, using techniques that draw on the scientific method. Realism in the social sciences asserts that social phenomena exist *independently* of our perceptions of reality, and that it is both possible and necessary to establish theoretical principles that approximate general social truths. This perspective on social research tends to be equated with quantitative methodologies, which are widely perceived as more rigorous and objective. However, realism is independent of methodological selection; many social scientists argue that quantitative techniques are as prone to subjectivity as qualitative ones, and that rigorous qualitative research techniques are equally suited to identifying and analyzing social phenomena.

[4] Some would argue that all scientific research endeavors, even those that aspire to the most rigorously objective knowledge, are inherently prone to subjectivity. That, however, opens up a can of epistemological and historical worms, and is a topic for another paper.

Other social scientists are less comfortable with the idea that social phenomena can be apprehended independently of our perception and conceptualization. They are reluctant to assert that objectivity and quantification necessarily enable social scientists to transcend the social worlds in which they are embedded. They point out that social scientists are themselves social actors who occupy particular positions in the social world. Not only do social scientists engage the people they are studying, but our social, cultural, linguistic, and psychological positioning necessarily shapes the lens through which we perceive, apprehend, and interpret what we study. In other words, social science is reflexive. Moreover, as some discussants pointed out, human behavior is reflexive: human beings interpret the world, act on those interpretations, then perceive and interpret outcomes; we then these interpretations of the world back into their behavior. This brings a recursivity and complexity to social dynamics that is absent in the physical world, because – as Bevir's paper makes clear – so-called "natural" (physical) phenomena are not responding to their ongoing interpretations of context. Representing reflexivity in computational modeling and simulation projects is a significant research challenge for computational social scientists.

3.3.1. Communicating Diversity to Consumers of Models and Simulations

Workshop participants agreed that neither decision nor policy-makers are likely to be aware of *epistemological*[5] debates unless they have formal training in the social sciences. In fact, many of the non-social scientists in the workshop commented that social scientists were discussing unfamiliar concepts and language. However, insofar as computational social modeling and simulation is a form of social science, we also agreed that funders and consumers of these technologies should at least be aware of such debates.

For one thing, this awareness helps explain the disunity of the social sciences (what one participant colorfully called their "jumbledness") as more than disorganization or hyper-specialization. After all, if disunity is an organizational problem, then it can be fixed if social scientists could identify and commit to a more coherent set of theories and practices. However, the social sciences are a heterogeneous set of theories, practices, and forms of knowledge with many different epistemological commitments. Just because all social sciences are "social" does not mean that they are necessarily compatible with each other. Cultural anthropologists, for example, often pursue a dialogic approach to research, co-creating narrative descriptions of social realities in concert with the people embedded in them. These dialogic approaches produce richly descriptive accounts of a community's social reality, but the idea that social scientists should collaborate with research "subjects" violates some researchers' beliefs about distance and research objectivity. Moreover, it is difficult to reduce these narratives into computationally tractable abstractions; doing may so rob them of the richness that makes them valuable.

Secondly, one's perspective on the nature of social science knowledge can influence the selection of modeling and simulation methodology, the collection of data, and the process through which models are developed and evaluated. For one thing, all models imply some degree of abstraction, which entails choices about what to include and what to filter out. As one discussant put it, all

[5] "Epistemology" was another term that the workshop participants flagged as potentially jargon-ish for decision and policy-makers, particularly those without graduate training in the sciences or social sciences. Epistemology is the branch of philosophy that addresses, literally, *how* we know what we know: the processes through which we seek knowledge establish facts, and justify truths about the world. Because epistemology provides the philosophical foundation for our assertions of truth and reliable knowledge, the fiercest debates among researchers often revolve around epistemological disagreements.

modeling efforts entail a "logic of selection" that necessarily influences simulation outputs, and that may or may not be articulated to decision makers (this is theme that Hemez addressed in his paper as well). One's logic of selection is likely influenced by one's underlying epistemological commitments, in addition to disciplinary-specific training; but - importantly - logics of selection can be assessed and evaluated when they are clearly stated and documented.

Thirdly, epistemological commitments influence how researchers approach the modeling process. Many modeling and simulation projects entail only minimal contact with the target communities or social groups under study (as when secondary datasets are used); or may not specify any particular social group at all but describe a general process instead. However, these are not the only ways to approach social science modeling and simulation. For example, two of the participants described collaborative modeling and simulation approaches, in which members of the communities under study become participatory stakeholders in the creation of the simulation, providing data and suggesting ways of applying findings[6]. These participatory modeling projects are similar to what Albro's paper describes as "joint co-creations" of ethnographic accounts. Similarly, several of the participants with experience researching user communities emphasized that incorporating users into the development of technologies may be critical if they are to understand the assumptions embedded in a simulation and apply them appropriately.

3.4. Computational Models of Culture

Military and intelligence decision makers are increasingly recognizing the importance of understanding culture as an important parameter in a range of problems, from religious fundamentalism to insurgency dynamics. This interest extends to creating computational models of culture (e.g., the Human, Social, Cultural and Behavioral Modeling program, which resides in the U.S. Office of the Secretary of Defense). However, the idea of developing computational models of "culture" is unfamiliar to many anthropologists, for whom computational modeling and simulation is not a mainstream methodology. Because cultural anthropology tends to produce narrative accounts ("ethnographies") of research findings, how to represent these narratives in computationally tractable format, without losing the contextual information embedded in narratives, is a difficult question.

Albro opened his paper by arguing that the U.S. Department of Defense's commitments to computational modeling and simulation are shape DoD approaches to studying and influencing culture - rather than cultural research demanding computational methodologies. For example, computational modeling and simulation projects typically demand data in particular structures and forms; in doing so, they set methodological agendas that are likely to exclude much of the data and information that cultural anthropologists develop. Computing implies an emphasis on code-able variables at the expense of narrative descriptions. Albro's reading of the metaphors that underlie computational social simulations in institutions like the U.S. Department of Defense emphasize concepts like *systems*, *structure*, and *functionalism* that emphasize the coherence and stability of culture, and/or that treat "culture" as an underlying mental state that shapes behavior. In contrast, many cultural anthropologists understand culture to be a kind of public discourse in which members of a community display, debate, re-interpret, and extend notions of collective identity.

---

[6]See, for example, Olivier Barreteau, 2003, "The Joint Use of Role-Playing games and Models Regarding Negotiation Processes: Characterization of Associations," *Journal of Artificial Societies and Social Simulation* vol. 6, no. 2 http://jasss.soc.surrey.ac.uk/6/2/3.html.

The group's discussion of Albro's paper once again highlighted differences in how social scientists approach the challenge of culture in their research. For example, several of the discussants pointed out that some cultural anthropology does incorporate quantitative methods and modeling techniques. Although many cultural anthropologists rely on qualitative ethnographic methods to construct narrative accounts of culture, there is quantitative tradition in cultural anthropology that incorporates both qualitative and statistical techniques. Another commented that Albro's paper represents one perspective on culture: that it is as a set of social understandings that are contradictory, loosely integrated, thinly coherent and contested. However, other anthropologists and sociologists describe culture as shared understandings and consensus worldviews that are more evenly distributed across members of a social group. Whether either formulation is reducible to a computational model was a matter of some discussion in the group.

Reflecting on Albro's paper, another participant commented that computational social science is not as strongly rooted in the social sciences as it is grounded in computer science, artificial intelligence, and operations research; it is perhaps not surprising that these practitioners have developed theories and methods that draw on their respective fields' root metaphors. She also emphasized that the U.S. Department of Defense openly seeks to influence social behavior, and metaphors like "system" usefully imply the possibility of intervention. However, while such metaphors open the door to computational techniques, they may also lead decision makers to exclude significant sources of information, such as qualitative data sources or narrative accounts, that do not fit well with dominant metaphors or analytic approaches. This could lead decision makers to adhere to preconceptions that are perhaps not well supported in the work of other researchers.

3.5. Prediction and Simulation Credibility

Francois Hemez' paper also raised questions about the epistemology of modeling and simulation, but from the perspective of computational physics and engineering. McNamara introduced Hemez' paper by observing that it is easy to assume that computational modeling and simulation of physical systems is a relatively straightforward process, compared to social modeling, given that the physical sciences are characterized by relatively stable theories that are mathematically expressible and enjoy high disciplinary consensus. However, all models are abstractions, which means that they necessarily simplify complex phenomena. In his work at Los Alamos, Hemez has considered and specified what it means to say that a model/simulation of a physical system is credibly predictive. McNamara and Trucano asked him to extend his thinking to computational models of social phenomena, which he very graciously agreed to do.

Hemez explained that as the nuclear weapons laboratories expanded their use of computational physics and engineering simulations in the wake of the Comprehensive Test Ban Treaty, the impact of these tradeoffs on the credibility of predictive simulations became more important, which led him to develop a series of arguments about what is required to demonstrate predictability across a class of models. Hemez's paper is an analysis of "predictability" in modeling and simulation when models are being used to augment lack-of-knowledge about phenomena of interest, and therefore represents a much more general discourse about modeling and simulation under conditions of uncertainty.

Computational modeling and simulation technologies enable researchers to instantiate and leverage knowledge toward phenomena that are not fully understood. However, because modeling and

simulation necessarily involves abstraction, the trade-offs required can affect whether or not a simulation is credibly predictive. Hemez's paper describes models as a kind of filter that resolves some phenomena more sharply at the expense of others. Some phenomena will be resolved in great detail, while others will be embedded in the overall model in the form of assumptions. Assessing the impact of these less-resolved assumptions on model-based predictions is critical if decision makers are to weigh results from models appropriately in their assessments. In his paper, Hemez laid out a set of concepts and a framework for examining predictability and assessing confidence in predictions.

In discussing the paper, one of the participants said that he could see the issues Hemez raised in the context of computational physics could be applicable in computational social science. He suggested, however, that some of the principles in would need to be recast for the social sciences, because the phenomena under study are even more unruly than the complex physics and engineering problems that Hemez describes. For example, first principles are rare in social science, and social scientists do not have classes of "equivalent" models that lend themselves to the kind of analysis that Hemez described in his paper.

Another participant commented that computational models and simulations in the weapons laboratories are part of a highly connected set of ongoing research activities. Computational social models and simulations are less embedded in complementary research and data collection activities. Moreover, this participant noted that Hemez linked predictability to concepts like robustness, fidelity, and consistency; for this participant, Hemez' approach to studying prediction was quite different than the "Nostradamus" goal of prediction that seems to be implied in many applied computational social science projects.

Several of the participants, in contrast, argued that physicists and engineers in the nuclear weapons programs actually do struggle with problems that are quite similar to the ones facing social scientists. In the weapons programs, for example, physicists and engineers are attempting to understand and predict phenomena that are outside their span of control. Reproducibility and control are often impossible for complex physics processes such as those that occur in an underground nuclear test. Moreover, the logic-of-selection problem is also in physics and engineering modeling for nuclear weapons certification. The destructive nature of weapons behavior means that scientists have to rely on indirect measurements; they have to infer meaning from what is gathered in the field, and these inferences may be based on the same principles that are being examined. Moreover, models set conditions for evaluating experimental data. At least in the challenges of modeling and simulation, he argued, the differences between the hard and soft sciences are perhaps less firmly delineated than one might assume.

Moreover, codes present their own complications. As physicists develop larger computation models, software engineering becomes increasingly important because how software is built affects how the model behaves. Models themselves can introduce non-intuitive behaviors, particularly in the case of very high-fidelity computational physics and engineering. Weapons programs codes as highly complex, with hundreds of thousands, even millions of lines of code. They can easily take on a life of their own: for example, when a code simulates a process that is not physically possible, this raises the question of how the code followed that particular path. Surprises are not unusual; researchers might *assume* that a code is coupling physics in a particular way, but later discover that the code did something different. Moreover, evaluating these codes is organizationally challenging. In the early days of computational science, codes were written by a single person who knew about all

the code, and who was likely to be the only user of the code. The codes are now so complex that no single person can know everything. Instead, parts of codes are the responsibility of local experts who know their part very well. However, because parts of the codes interact with each other, issues are difficult to identify and analyze: they are never in a single piece of code, but in the interactions among them. It is difficult to diagnose such complex software products.

3.6. Models, Predictive and Otherwise

Hemez' paper opened a long discussion about the interrelated topics of prediction, verification and validation, and the physical and social sciences. However, a more fundamental question soon emerged: Do modeling and simulation projects always entail predictive goals, or are there other, non-predictive roles for modeling and simulation projects? This is not just a problem for national security, but for any domain in which modeling and simulation is used to support decision making. This topic provoked a long, sometimes vehement exchange about "prediction:" if orientation toward the future always entails prediction; if prediction is possible when knowledge is incomplete; and whether different kinds of prediction exist.

One seminar participant asked if model developers, users, and other stakeholders always expect models to predict, or do they treat them as instruments that enable understanding of a phenomenon, as models are often used in the natural sciences (though one seminar discussant argued that "understanding" may itself entail prediction)? This is a key question that the modeling and simulation community must address in its interactions with the user community. Funders of computational modeling and simulation projects may emphasize prediction as a significant goal for the research they support. It is necessary to challenge the culture and ethos of the client community so that they can use and interpret models in a more sensible way.

Another participant with experience in military modeling projects argued that the process of modeling has important uses beyond forecasting or prediction. If organizations are using simulations to build understanding about a problem space, identify ranges of possible outcomes, or to support self-assessment and planning, are they really doing prediction? Formalizing implicit beliefs and assumptions can help people recognize the incompleteness of their knowledge. Participatory modeling and exploratory modeling are approaches that enable people to understandings about the world, and to identify ranges of possible outcomes for events and actions. These are worthy and legitimate reasons to do modeling beyond "prediction."

In contrast, another participant argued that possibility is actually a form of prediction, because any scenario that is identified as possible involves a cluster of features, about which a claim of relationships and order is being made. Others disagreed, saying that prediction indicate a specific and narrow claim about the future. One of the social scientists with experience in modeling and simulation argued that prediction in the social sciences may indeed be a different kind of prediction, one that may less expressible in quantitatively precise terms, but may permit the exploration of counterfactuals. In many cases, it may be impossible to do more than qualitative prediction, or to identify boundary conditions for possible outcome spaces. In many cases, it may be impossible to do more than qualitative prediction, or to identify boundary conditions for possible outcome spaces. As another participant emphasized, however, even qualitative "predictions" can and should be subject to evaluation.

In discussing the limitations of predictive modeling, one of the participants described an ecological experiment looking at intertidal food webs. In the experiment, the top predator was removed from the food web and it was predicted that the prey items of this predator would directly benefit from the removal (i.e., each would flourish with no predatory pressures). However, this ultimately caused an explosion in the population of a particular barnacle due to its ability to out compete all other intertidal species for space. This outcome that was completely unforeseen in the modeling process. Complex systems have latent and emergent properties that can make prediction extremely difficult. These observations recalled Hemez' points about the limitations of modeling when knowledge of the domain represented in the model is incomplete. Understanding the sensitivity of model outputs to lack-of-knowledge and embedded assumptions are critically important if modeling and simulation results are to be used in decision making.

As McNamara and Trucano pointed out in their provocation paper, the CSS literature uses a variety of synonyms for "prediction," such as forecasting and anticipation, in trying to differentiate among different types of prediction. What terms mean in practice is rarely defined and likely not well understood, which leaves a lot of room for claims that can neither be evaluated nor substantiated. Moreover, the level of rigor associated with these different kinds of "prediction" can vary tremendously.

3.7. Verification and Validation[7]

Turnley's paper also explored the problem of prediction, in relation to verification and validation. In the physical sciences, "prediction" is often tied to "validation," in the sense that validation requires the comparison of a model's prediction against an observed, real-world effect. If we assume that some uses of models do *not* entail "prediction," then perhaps validation approaches grounded in predictive fidelity are setting unreasonable standards for assessing their external correctness. For example, she pointed to face validity as an alternative way to address the credibility of a model.

One participant commented that specifying many different types of validation could lead to confusion on the part of stakeholders. Another participant commented that Turnley's argument could be read as justifying avoidance of rigorous validation, which involves the identification of referents that support systematic assessment of the model's relationship to the external world. He acknowledged that identifying referents is difficult in the social sciences, and observed that validation presents pragmatic challenges when time and money are short. A third argued that validation is a process that begins with specifying the intended use and application of a model; once the stakeholder community agrees on that usage, appropriate techniques to assess external correctness of the model should derive from that. Face validity may be fine if that is consistent with how the model will be used; however, he pointed out that techniques like face validation rely on

[7] Verification and validation (V&V) are processes that assess modeling and simulation technologies for internal correctness (verification), and external correspondence to real-world phenomena of interest (validation). There is an enormous body of literature dating back to the 1970s that addresses methods, techniques, tools, and challenges for V&V. Most of this research has been done in fields like computational engineering, artificial intelligence, and operations research. However, in the computational social science community, there is an emerging body of literature addressing the challenges of verifying and validating computational social science models and simulations. For more information, see McNamara and Trucano paper, Appendix A, page 9.

experts, which raises the question of who the experts are, how they are selected, and how their expertise vis-à-vis the model is evaluated.

One of the participants, a researcher with extensive experience in government agencies, agreed that model validation is extremely important, but questioned the extent to which frameworks and concepts from arenas like weapons physics are applicable to computational social models and simulations. The rapid evolution of social dynamics, and of national interests in those dynamics, presents significant challenges for model construction and evaluation – particularly verification and validation, which can be time and resource intensive. The problems that analysts deal with evolve so quickly that very detailed, contextually specific models and simulations will have a short shelf life. Because models have such short self-lives, and the decision cycles are so short, V&V is challenging.

Where to draw the line between rigorous assessment and pragmatic judgment in CSS applications, given the operational challenges of gathering and analyzing social science data, is a difficult challenge. Organizations struggle with this problem, and it is unclear what resources are being provided to support verification and validation research vis-à-vis computational social modeling and simulation programs in large government agencies. It is often difficult to get institutions to put the time, financial, and personnel resources into verification and validation activities.

3.8. Usability and Utility in Computational Social Science

Verification and validation have an analogous relationship to the twin concepts of *usability* and *utility*. Usability testing assesses whether a software product is constructed so that people can operate it. However, usability says nothing about the relevance of the software product to the work that it is intended to support. *Utility* assesses whether the software's functions map to the real-world tasks and goals of its intended human users. A related concept is adoptability: the degree to which a technology will be picked up for use across the user community for whom it is intended. These were major themes in the workshop and were explicitly addressed in both Huxtable and Scholtz's papers. Both authors asserted that usability, utility, and adoptability constitute difficult issues for CSS technologies being leveraged in the national security community.

Scholtz's paper discussed the problem of transforming computational modeling and simulation projects into analytical tools that people will find usable, useful, and trustworthy. CSS tools raise particularly vexing challenges in this regard, as analysts may not fully understand how models are generating solutions. They are unlikely to adopt new technologies if they do not understand how they work, and if they do not trust them. Studies of technology trust provide varying perspectives on this issue. Some suggest that analysts tend not to trust technologies that have been developed by outsiders who do not appreciate the analytic process. Others assert that "black box" technologies are fine, as long as the users understand the inputs and outputs. In the case of computational social modeling and simulation technologies, however, users may be doing parameter studies (for example, to assess ranges of possible outcomes given variations in initial conditions), that require some deep familiarity with the model and the phenomena it covers. When analysts do not understand the intricacies and subtleties of a problem, they may have less confidence in their assessments.

One way of addressing this problem is to ensure that users of computational social modeling and simulation technologies are involved from the earliest stages of the development process. However, who the "users" are is not clear. In discussing her paper, Scholtz identified at least three types of

stakeholders in these projects: analysts who run the simulations, policy and decision makers who consume information; and the technology and/or subject matter experts performing the research and development. The multiple user communities make "utility" design and evaluation a challenging problem. Participatory modeling is one way of supporting deep familiarity with models and domains, but it is difficult to do this in the intelligence community because of access issues and constraints on analysts' time. She suggested a range of alternative methodologies for gathering data about analyst interactions with information, not only to develop tools that are more adoptable but to identify ways of teaching the users about how to properly apply the technology in a given problem space.

Several of the anthropologists pointed out that ethnography can be useful in this regard; for example, organizational and development ethnography provide techniques to help technology developers, funders, and advocates understand how new tools disrupt workflow or shift access and control over information resources. However, others observed that user-oriented design techniques are unlikely to solve the adoption challenge. Instead, government managers may have to play forcing function to get analysts to adopt new tools. Workforce training and developing will also be important in encouraging analysts to adopt new technologies; for example, getting an Army sergeant to use a computational simulation requires education, just as piloting an airplane requires skills, training, and experience. Modeling and simulation tools cannot in and of themselves "package" expertise for delivery to external user communities that may lack domain experience, modeling and simulation expertise, or both. For some kinds of modeling there may actually be requirements (education, training, experience) placed on *users*, as opposed to identifying user requirements for software.

In discussing his paper, Huxtable agreed with Scholtz' observations about multiple stakeholders in modeling projects – researchers, technology developers, users – each of which have distinct interests. He described a demand among analysts and decision makers for techniques to characterize complicated social, political, and historical problems. However, despite substantial government investments in a range of modeling projects, Huxtable believes that most modeling and simulation projects are more "proof of concept" than accepted tool. He related this to several factors: first, government contracting rules restrict the kinds of interactions that can take place between technology developers and end-users. Second, interacting with academic experts is challenging. Unlike their academic counterparts, military and intelligence analysts worry about how to intervene and effect change. They are charged with implementing policy made at higher levels of government, are answerable to these higher levels of government, and often work on tight timelines and budgets. Third, academia does not generally reward transdicisplinary or applied research, which lessens academic motivations to partner across disciplinary and organizational communities to develop working tools. And finally, many researchers do not want to get security clearances. This makes hard for them to interact with the users, and also creates duplication of effort, since academics on the outside are often not aware of complementary classified work.

Huxtable suggested that relatively little attention has been paid to the more basic issue of how social science research *methods* might be used in military and intelligence analysis. It is deceptively easy to find problems that fit computational tools. Identifying and integrating analytic and research techniques first would help developers identify computational requirements as (and if) they emerge. Secondly, he wished that there were more flexible tools that could support a range of analytic approaches, such as libraries of agents or systems dynamics techniques that can be quickly applied as

appropriate. These tools should be adequately evaluated and well-documented, so that users know what they are using when they pull such tools off the shelf.

An even bigger problem, however, is assessing whether or not modeling and simulation technologies improve decision-making. Even demonstrably "better" models will not necessarily improve decision-making, as people in the military and intelligence communities face a range of constraints beyond the tools available to them. One participant noted that decision makers rarely base their assessments solely on the results of a model; in fact, it is difficult to assess how and to what degree a model contributes to that decision, much less to an outcome. These issues complicate how we evaluate the impact of new technologies, including CSS models and simulations, on decision outcomes.

3.9. Ethics, Modeling, and Decision-making

Closely related to the topic of users and usability is the issue of ethics. The problem of "ethics" in "computational social modeling and simulation" is not a mainstream topic of conversation in the CSS community. In her paper, Lucy Resnyansky observed that ethical discourse around computation tends to focus on issues of privacy, confidentiality, and civil liberties. Some of these issues may pertain to computational social science as well (as when researchers use communications datasets to study social patterns), but computational social modeling and simulation, particularly in its applied form, probably raises its own quandaries.

Resnyansky noted that computational social science represents the intersection of a number of fields, including social science, information technologies, and computational modeling and simulation, each of which have their own ethical challenges and underpinnings. For example, companies that sell enterprise software systems are deploying large data collection, tracking, management, and analysis tools that can have a profound impact on management effectiveness and decision-making. Similar challenges may be present in applied computational social science, where software models are intended to affect decision-making and outcomes in areas where public trust, resources, and even human lives may be at stake. Issues of responsibility and accountability are paramount, but difficult to locate. In addition, when new tools are introduced into a workplace, they may impact organizational workflows in surprising ways. As several workshop participants pointed out, little is known about the diversity of practices in intelligence and defense analysis work, and work practices vary tremendously across contexts.

Resnyansky allowed that computational methodologies offer tremendous advantages for social research; for example, they allow researchers to replicate and study problems for which real-world experimentation would be logistically difficult and/or ethically reprehensible, such as the spread of a virus through an urban population. However, developing models of complex social domains should require some research familiarity with those domains, and this is not always present in computational social modeling and simulation efforts. In her work with modeling teams, Resnyansky had noticed that terms like *social*, *human*, *cultural,* and *behavior* are not always well defined. She spoke of the intersection of computational social science and national security as a merging of cultures that are not familiar with each others' conceptual assumptions. This cultural divide is complicated by the fact that computational social modeling and simulation in national security is dominated by engineers, computer scientists, and physicists who lack theoretical and methodological training in the social sciences, and who are also unfamiliar with social science research ethics. As several of the

participants observed during the workshop, computational social modeling and simulation projects are often approached as modeling problems first, and as social science problems second. Several of the workshop participants relayed examples of non-social scientists - including computer scientists, physicists, and engineers - developing modeling and simulation technologies for social phenomena, often without expertise in the relevant social science domain.

Resnyansky's paper raised another important question: what responsibility do the developers of computational tools have for the effects of their technologies downstream? In most national security environments, the "user community" is not unitary; there are people who run models, analysts who incorporate model outputs into their analytic tradecraft, policy and decision makers who consume the information products that analysts develop, decision makers who take action in the field based on what analysts tell them. This raised the issue of whether computational social simulation is being oversold to national security decision makers. Resnyansky told the workshop participants that she had reviewed the popular discourse on modeling and simulation, which she described as "promotional literature." She was struck by the way the literature presented techniques and methods as "good," without much discussion of limitations or risks. Huxtable expressed particular frustration with the mantra of "situational awareness" that gets attached to computational modeling and simulation. The complicated and subtle nature of computational social simulation, and the challenges of transitioning that knowledge to users who likely lack similar expertise in the subject matter and/or the modeling approach, make claims of computationally-supported situational awareness difficult to believe. It is a truth in advertising problem, he said.

Responding to Resnyansky, one of the participants commented that the ways in which modeling and simulation technologies are developed might shape the ethical relationship between developers and end-users. If models are artifacts that are "thrown over the fence" to users, that may entail a different set of responsibilities than if the developers are offering to support a process of modeling in the context of stakeholder discussions. In the latter case, the modeler is providing an analytic service that is couched in a different set of client-developer expectations and products, and may even be considered part of the analytic process.

Lastly, the workshop participants agreed that modeling proponents and technology developers need to address issues of accountability and responsibility, though whether this is recognized in the CSS community is questionable. A workshop participant described feeling troubled about a perceived lack of critical concern about the kind and quality of knowledge that modeling and simulation introduces into decision-making. It is important to understand how the norms that govern democratic societies, the norms of governance, intersect with the decision making processes that we expect from government, and the role of supporting technologies within those.

3.10. Data, Error, and Social Networks

As the final paper in the workshop, Jeffrey Johnson's presentation on social network analysis was unique because it focused on particular kind of modeling and simulation methodology. However, his paper also illustrated most of the themes that the workshop participants had discussed during the previous two days.

McNamara and Trucano felt it was important to get a social network researcher involved in the workshop, because there is so much emphasis on the potential for social network analysis for

national security decision-making and confusion over what social network analysis actually entails. Data is a critical driver for this enthusiasm: the national security community is faced with an explosion in data sources that lend themselves to representation in node-and-arc graphs, such as communications records. Indeed, link analysis is commonly used in the intelligence community, and involves using node-and-arc diagrams to represent a mental model of relationships among entities in a dataset.

However, Johnson emphasized that social network analysis is not the same thing as link analysis. Indeed, social network analysis is different technique altogether, because it involves using mathematical analysis to characterize the structural attributes of a community, and to assess the roles that individual actors play in a community or group. This kind of analysis can be powerfully revealing of a group's organizational properties, but it is also very sensitive to the quality and completeness of data, in ways that may not be apparent to naïve users. Even small amounts of missing data can dramatically impact the conclusions one draws about the posited roles of particular actors in networks. In addition, assumptions about what data means can also lead to erroneous conclusions about the nature of a network. Another participant agreed, and pointed out that the problem is not just whether or not data are missing, but what data signify. Many of the basic social network analysis techniques assume that ties of the same type are formally equivalent, but this glosses over significant differences in the content of the relationship or the intent of the actors on a network.

As a very basic example, Johnson described working with military analysts using communications records to map social networks. These analysts had not considered that the communication node – say, a telephone number – provided only minimal information about the act of communication: who was making a call, who was receiving a call, and what was exchanged. More subtly, building social networks from secondary data sources, such as intelligence reports or newspaper articles, is a common practice in the intelligence community. However, this raises some very difficult questions about the extent to which these secondary information sources can be used to construct valid and complete datasets. Johnson identified multiple sources of uncertainty: erroneous reporting, incomplete information, deception, none of which may be immediately apparent to the analyst.

4. RECOMMENDATIONS

The workshop convened in Santa Fe explored a broad range of issues related to computational social modeling and simulation, from the impact of new computational techniques on social science research to the practicalities of tool adoption among intelligence and military analysts.

In the wake of the workshop, we (McNamara and Trucano) reviewed the insights, commentaries, and tremendously good work of the participants who brought their ideas and papers to Santa Fe, and distilled the following recommendations. **These recommendations are based on our reading of the papers and review of the workshop discussions, and are not necessarily representative of all the participants' views in the workshop.** In fact, we believe that our workshop participants could probably each write their own sets of well-informed recommendations with nuances that we have missed. Also, this is not an exhaustive list of the recommendations we could have made, nor did we cover all the points and nuances that could have been raised in the explanatory text that follows each. Instead, we attempted to identify the most important overarching themes, and to assemble these into a set of interrelated conceptual starting points for making sense of these enormously complicated technologies. We hope the following five recommendations will give people both inside and outside the computational social modeling and simulation community with a credible starting point to develop well-informed and judicious questions about the benefits and risks of using models and simulations in high-consequence decision making environments.

RECOMMENDATION ONE: Design, implement, and assess computational social science projects as hybrid, interdisciplinary research and development efforts.

We assert that computational social modeling and simulation projects should be analyzed and studied as the interdisciplinary processes and technologies that they are. Three domains in particular are germane to any assessment of computational social science models and simulations: *social science*, *computational science*, and *decision support and analysis software*. When computational social science technologies are applied in decision-making, these domains intersect, as depicted in the diagram below:

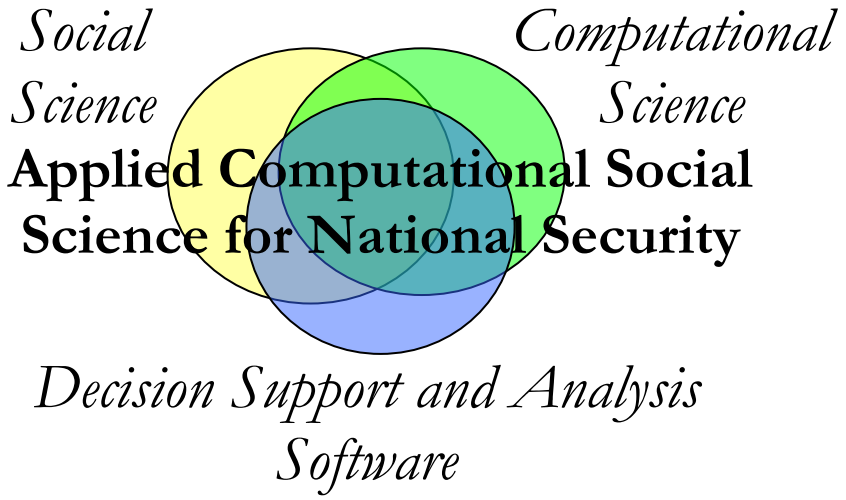


Figure 2: Applied CSS as Interdisciplinary Domain

Only an interdisciplinary framework that addresses the multiple domains of research and practice that pertain to these technologies has any hope of ensuring complete assessments of their quality and correctness. The next four recommendations derive from this intersection, and pertain to the design, development, and evaluation of computational social science projects in decision-making environments. Although we have focused on national security decision-making, we see no reason why the recommendations we present are unique to national security. Instead, they are applicable to any domain in which computational social modeling and simulation technologies are being developed to augment human sensemaking for the purpose of supporting decisions, planning, and action.

RECOMMENDATION TWO: Evaluate computational social modeling and simulation projects as a form of *social* science.

Computational social science is a first and foremost a form of social science. Modeling and simulation is a field of methodological research for studying social, cultural, and behavioral phenomena. Its application is only as "scientific" as the research design in which it is embedded.

Reframing computational modeling and simulation as a methodology enables national security decision makers to compare it to other methodologies and approaches for addressing the problems they care about. Indeed, one question that needs to be asked more frequently and consistently is, "Why does this problem require computational modeling and simulation?" As intriguing and compelling as modeling may be, not all national security problems that touch on culture, society, or human behavior are necessarily amenable to modeling. Nor is computational modeling and simulation always the most accurate, robust, fastest, least expensive, or most portable way to develop and deploy knowledge of a domain. The intense focus on computational simulation as a superior methodology for studying social dynamics comes at the expense of careful consideration of other forms of social science research and practice, and perhaps even existing knowledge and expertise. Social scientists have developed a range of methods and approaches for studying human social, cultural, and behavioral phenomena that concern national security decision makers. These need to be tapped as well; indeed, there are probably significant gains to be realized from coupling modeling and simulation with more traditional research approaches in the social sciences.

Secondly, the intense focus on modeling and simulation means that many computational social science projects are dominated by people with mathematics, physics, and engineering backgrounds, at the expense of domain and methodological expertise from the social sciences. This creates risks for decision makers, because models are abstractions that always entail a particular selection logic. Modelers without domain knowledge may not appreciate the significance of what they are including, the relationships they are specifying, or – most importantly – what they are leaving out. This is particularly true if the model generates simulation outputs that *seem* to replicate real-world phenomena; replication does not necessarily entail validity, explanation, or prediction. There are many wrong ways to get the right answer. Ensuring that relevant domain, theory, and methodological experts are involved in the modeling and simulation exercise is one way of addressing this problem, although we realize that the selection of a domain expert raises its own challenges.

Decision- and policy-makers who are attempting to make sense of a particular computational social modeling and simulation effort should ask questions about scientific validity of the theoretical and conceptual framework that underpins the model. They should ask how the modelers identified and assessed existing research on the domain, how they connect their work to pertinent questions in the social sciences, and to explain how they identified questions that the model is seeking to answer. Modelers should also be able to specify and justify the role of modeling in their research design. They should speak credibly to the challenging problems of data: how they obtained data, if these datasets are primary or secondary data, who collected the data, how, and why; to justify why data are relevant to the problem, and to discuss pertinent assumptions, limitations, and sources of error in the data. They should be able to explain why particular modeling and simulation approaches are necessary for examining this domain, and to articulate the basis for claims to the credibility of their simulation outputs.

These are the kinds of questions that most social scientists are accustomed to answering in project proposals and reviews. Insofar as they are doing social science, computational modeling and simulation practitioners should be prepared address them as well. In addition, reviews of computational social modeling and simulation efforts should seek to involve social scientists with experience in selecting and applying different research methods, including computational modeling and simulation techniques. Reviewers should include researchers who are *not* receiving national security funding, to minimize possible resource bias.

RECOMMENDATION THREE: Evaluate computational social models and simulations as a form of *computational* science.

As important as it is to couch computational social modeling and simulation projects in the social sciences, it does not go far enough. Social science does not ask questions about algorithms or underlying mathematical issues, nor does it point to issues of software implementation, testing, and performance. However, the field of computational science does, and framing computational social science as a form of computational science raises a second set of issues for evaluating these projects.

We recognize that computational social science deals with complex, often inaccessible phenomena and processes, such as the recursive relationships among perception, interpretation, and action in human dynamics. While this creates considerable validation challenges, it does not mean that principles from other fields of computational science are irrelevant to computational social science. On the contrary, we believe that computational science offers decision makers with mature conceptual frameworks for evaluating computational social modeling and simulation efforts as a form of computational practice.

Computational models and simulations are complex objects of study in their own right, particularly when they are being used to predict or forecast phenomena for which we have limited understanding. Whether or not a computational artifact is properly constructed has tremendous impact on the credibility and trustworthiness of its outputs. Computational science offers a range of concepts and methods for studying and evaluating computational models and simulations as computational artifacts. Many of these concepts and methods are very basic and domain agnostic, at least at the conceptual level: for example, *verification* is about evaluating whether or not the software is correctly implementing a conceptual model. It evaluates the correctness of software, not how the

model deals with subject matter. Similarly, good software engineering practices, well-designed testing processes, and ongoing documentation can ensure that modeling and simulation projects are not prone to inaccuracies and errors that stem from poorly written codes. Moreover, computational scientists who are using modeling and simulation methodologies to assess and support decisions about real-world outcomes have developed theoretically sophisticated frameworks for understanding the limitations of modeling as a source of predictive information under conditions of uncertainty. For example, the closely related field of uncertainty quantification specifies techniques for addressing different forms of uncertainty in modeling processes, and for examining how uncertainty should be treated in decision-making.

Decision makers should ask about software engineering practices, documentation, and testing, as well as how the modeling team is evaluating the correctness and performance of the software it has written; how sources of error in the code are detected and mitigated. At a deeper level, the experience of other fields may provide valuable input to the challenges and issues that arise when computational modeling and simulation outputs are being used as a source of predictive information in decision-making, when knowledge is lacking.

RECOMMENDATION FOUR: Evaluate computational social models and simulations as *decision support tools* for individual and organizational use communities.

It should not be taken for granted that a modeling and simulation technology is a decision-support tool. Tools and technologies are different things. To say that technology is a *tool* is to assert that it can be employed by human beings to effect some change in their understanding of the world, or in the world around them. Tools afford efficacious human action: they are *usable and useful*. They fit well into contexts of use: they are *adoptable*. Technologies that meet these criteria are likely to be used.

In short, computational social modeling and simulation technologies raise complicated questions about the relationships among creation, use, and outcomes. Even modeling and simulation technologies that have undergone rigorous verification and validation may not meet the demands of client, user, or information consumers in government environments. Relationships and responsibilities must be considered before the tool development begins: For example, who decides the goals of a model: the people responsible for making decisions, or the modeling experts? Are they one and the same, or is the modeling project distributed across multiple stakeholders? Who are the "users" and what is their relationship to the modeling project?

These questions can be very difficult to address in the fractionated environments that characterize government policy and decision-making. Moreover, the kinds of decisions that computational social modeling and simulation technologies are expected to support are complicated, ambiguous, often poorly understood, and rapidly evolving; and rarely is there a single point of decision-making, nor a unitary decision maker. Issues of context, users, consumers, and communication are critical if modeling and simulation technologies are to mature into tools.

Definitions of usability, utility and adoptability that come from the "seller" of the model are not acceptable. Setting standards for usability, utility, and adoptability is a task that belongs squarely in

the domain of the client or consumer. We assert that design processes for these technologies should involve some attention to the *intended areas of application* (this is necessary for validation *and* utility) and the *intended user communities*, if only to ensure that the resulting technologies are both usable and useful. A good start is to draw on the research in organizational ethnography to identify factors that influence usability, utility, and adoptability in decision analysis tools Moreover, modeling and simulation technologies may support insight into complicated social, cultural, and behavioral dynamics, but these technologies are not transparent. Models often embed assumptions and concepts whose implications and limitations are difficult to appreciate without domain expertise. Participatory approaches that involve users in the development process are more likely to produce educated consumers/users of both the tool *and* the information it generates.

We note as well that what means for models and simulations to improve analysis, decision-making, or provide insight, is rarely well specified. This makes evaluation of impact very difficult. Human-computer interaction, human factors, and cognitive psychology can be leveraged to develop studies that assess the impact of modeling and simulation technologies on how people assess and draw conclusions from complicated and ambiguous datasets. There is an extensive literature on the design and evaluation of information visualization and visual analytics tools, two related fields that dovetail nicely with computational social modeling and simulation (which often entails rich and colorful representations of human dynamics). These and other areas of literature that address human sensemaking and technology interactions can help technology developers provide valuable toolsets that demonstrably improve human understandings of complicated problem spaces.

RECOMMENDATION FIVE: Support interdisciplinary exchanges that enable computational social science researchers, developers, adopters, proponents, users, and stakeholders to learn how other fields analyze and evaluate models and simulations.

While we recognize that there are significant differences between the physical and the social sciences, we continue to assert that other fields' experiences with developing, deploying, interpreting and applying modeling and simulation technologies can help organizations understand how to develop models, interpret their outputs, combine simulation outputs with other forms of information, and assess the limitations of modeling and simulation technologies in decision-making. Models are not just tools for analysis; they are artifacts that require analysis if we are to understand how they function, how we can use them responsibly, and what their limitations are.

5. Conclusion

Computational social science models and simulations are hybrid, interdisciplinary technologies that bring computational methodologies to the study of social, cultural, and behavioral phenomena. Computational modeling and simulation methodologies are well-established in many other areas of science; their adoption among social scientists is a relatively new phenomenon. Many social science researchers are excited about computational modeling because it offers a range of techniques for exploring social, cultural, and behavioral dynamics that are difficult to systematically observe and document in the real world. Some of the most creative, promising research projects employ

modeling as part of a multi-method suite that includes real world observation and more traditional qualitative and quantitative research techniques. However, the very nature of the dynamics under consideration – geographically distributed, taking place over multiple time scales, involving human cognitive processing and intentionality – presents significant challenges for empirical data collection – hence the attraction of computational techniques.

Similar difficulties confront decision makers charged with analyzing and developing strategies for effecting outcomes in complex social, political, and cultural spaces. While U.S. government agencies have learned a tremendous amount about the national security challenges that face the country in the post 9/11 era, they are also still learning how to ask the right questions, and information and data about critical processes and events remain sparse and ambiguous. Decision-makers in places like the U.S. Department of Defense and the Intelligence Community are typically working in environments where timelines are tight, resources limited, outcomes uncertain, the organizational politics complicated, and consequences high. Just as researchers look to computational modeling and simulation to generate and explore ideas about the origins and evolution of social, cultural, and behavioral dynamics, so too are decision makers seeking technologies that can provide insight into these complex domains.

However, it is critical to remember that the use of modeling and simulation for research purposes is a different project than the use of modeling and simulation technologies for decision-making. In a research environment, errors in data collection, modeling, simulation, analysis and interpretation can undermine the validity of research findings. In government decision-making environments, the same errors may have significant, even existential consequences, particularly when kinetics are involved. Modeling and simulation methodologies are novel, dynamic, exciting, and attention-grabbing. However, they are not magical: they are analytical methodologies, prone to a range of errors and problems that are still not well understood. Decision and policy-makers would be wise to broaden investments beyond just models themselves, to encompass the complex challenges of developing, deploying, and evaluating computational technologies as sensemaking tools for human beings working in high-consequence decision environments.

6. List of Appendices

Appendix A: Abridged Version of McNamara-Trucano Paper
Appendix B: Workshop Agenda
Appendix C: Workshop Papers
Appendix D: Participants' Full Biographical Statements

APPENDICES

Appendix A: Abridged Version of McNamara-Trucano Paper

Challenges for Computational Social Science in National Security Decision-making: Prediction, Evaluation, and Users

Why Models Don't Forecast

Laura A. McNamara, PhD
Timothy Trucano, PhD
Sandia National Laboratories
Albuquerque, NM 87185
lamcnam@sandia.gov
SAND-2010-5203C

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Introduction

The sub-title of this paper, "Why Models Don't Forecast," has a deceptively simple answer: Models don't forecast because *people* forecast. Yet this statement has significant implications for computational social modeling and simulation in national security decision-making. Specifically, it points to the need for robust approaches to the problem of how people and organizations develop, deploy and use computational modeling and simulation technologies.

In the next twenty or so pages, we argue that the challenge of evaluating computational social modeling and simulation technologies extends far beyond verification and validation, and should include the relationship between a simulation technology and the people and organizations using it. This challenge of evaluation is not just one of usability and usefulness for technologies, but extends to the assessment of how new modeling and simulation technologies shape human and organizational judgment. The robust and systematic evaluation of organizational decision-making processes, and the role of computational modeling and simulation technologies therein, is a critical problem for the organizations who promote, fund, develop, and seek to use computational social science tools, methods, and techniques in high-consequence decision-making.

Computational Social Science in the Post 9/11 World

Computational social science is a diverse, interdisciplinary field of study whose practitioners include (but are not limited to) computer scientists, physicists, engineers, anthropologists, sociologists, physicists, and psychologists. Computational social modeling and simulation has lineages in computer science, mathematics, game theory, sociology, anthropology, artificial intelligence and psychology, dating back to the 1950s. However, the application of computational simulation to social phenomena exploded in the 1990s, due to a number of intellectual, social and technological trends. These included the popularization of complexity studies [1, 2]; the rapid spread

of personal computing throughout multiple facets of work and social life; the rise of electronic communications technologies, including the Internet, email, and cellular telephony [3-5]; the subsequent explosion of interest in social networks [6-10], and the development of object-oriented programming. Together, these generated new sources of data about social phenomena, democratized computational simulation for researchers, and opened the door for a creative explosion in modeling methodologies and techniques [11, 12].

Researchers in a range of fields see tremendous promise for computational social modeling and simulation as a technology for producing knowledge about human behavior and society. Modeling usefully supports development and refinement of hypothesized causal relationships across social systems, in ways that are difficult to achieve in the real world [13]. For example, agent models allow researchers to develop artificial societies in which "social scientists can observe emergent behaviors in terms of complex dynamic social interaction patterns among autonomous agents that represent real-world entities" [14]. Moreover, researchers can and do use simulated data instead of, or in addition to, real-world data [15]. Researchers in a range of fields are using these new modeling techniques to explore phenomena that are difficult to study in the real world because of ethical, temporal or geographical constraints; and to implement conceptual models or theoretical abstractions and simulate outcomes using the computer as a kind of "in silico" laboratory [16, 17].

Perhaps not surprisingly, a kind of revolutionary excitement and anticipation permeates much of the interdisciplinary literature on computational social science [18-20]. For example, David Levin, professor of public policy at Harvard's Kennedy School recently argued that, "social science will/should undergo a transformation over the next generation, driven by the availability of new data sources, as well as the computational power to analyze those data." [8] Many computational social scientists believe that we are on the brink of a computationally-driven paradigm shift that will

[8] http://www.iq.harvard.edu/blog/netgov/2009/02/paper_in_science_tomorrow_on_c.html

change social science permanently [17-20]. For example, political economist Joshua Epstein has argued that agent-based modeling and complexity thinking are driving a broader conceptual shift to an explanatory or *generative* social science in which the ability to computationally generate social phenomena becomes a standard for evaluating truth claims [17, 21].

A number of practitioners in computational social science not only see a promising future for computationally-enabled social research, but also believe that policy and decision makers would benefit from using computational modeling and simulation technologies to understand the complicated social, political, and economic events, and perhaps support the formation of more effective policies. For example, in the wake of the recent financial crisis, physicist J. Doyne Farmer and economist Duncan Foley argued in *Nature* that econometric and general equilibrium models are inadequate for understanding our complicated economic system; that agent-based models can help decision makers formulate better financial policies; and that an ambitious goal would be to create "agent-based economic model capable of making useful forecasts of the real economy" ([22]: 686). Similarly, Joshua Epstein opined that policy and decision makers would benefit from using agent-based modeling techniques to understand the dynamics of pandemic flu and make appropriate interventions [23].

This brings us to the issue of computational social science in national security policy and decision-making. It is worth noting that as the Cold War was coming to an end in the late 1980s and early 1990s, computational social science was experiencing explosive growth. This confluence perhaps explains why so many decision makers in federal departments and agencies are looking to computational social science to meet some of these new technological needs. In particular, the 9/11 attacks mark an important turning point in the relationship between the computational social science community and national security decision makers. The reader may recall how several observers working with open-source information (i.e., newspapers and Internet) developed retrospective (and

here we emphasize the word *retrospective*, since so much of the national security discussion in this regard is focused on forecasting) social network analyses that very clearly "connected the dots" among the attackers [24]. One highly publicized example came from organizational consultant Vladis Krebs, who spent weeks combing through newspapers to find information about the hijackers, piecing together a sociogram that mapped relationships among the participants. Krebs argued that the Qa'ida network was optimally structured to address competing demands of secrecy and operational efficiency; and pointed out that social network might be useful as a diagnostic tool to identify and interdict criminal activities. Soon after, Krebs was asked to brief intelligence experts on the analysis and detection of covert networks [25-27].

Of course, the idea that analysts should have been able to forecast the 9/11 events using signs that are retrospectively obvious is a case of hindsight bias [28, 29]. Moreover, the government's failure to interdict the 9/11 plot before the attacks involved multiple failures beyond simply connecting the proverbial dots, with or without a sociogram [30]. Nevertheless, analyses like Krebs' drew both popular and government attention to the idea that arcane research areas like graph theory, social network analysis, and agent-based modeling might be predictive, at a time when terrorism research was undergoing "explosive growth" as measured by publications, conferences, research centers, electronic databases, and funding channels [31]. Over the past decade, a number of computational social scientists have argued that modeling and simulation techniques are uniquely suited to understanding the dynamics of emerging threats, at a time when national security decision makers are urgently looking for new frameworks, data sources and technologies for making sense of the post 9/11 world [32-35]. Indeed, within the computational social science literature, there is a significant sub-category of post 9/11 academic and policy writing that examines how computational social modeling and simulation, particularly agent-based simulations in combination with social network analysis techniques, might enhance understanding of a wide range of national security

problems, including state stability, insurgency warfare, bioterrorism, flu pandemics, and terrorist network detection (see [25, 26, 32-63]; also [27, 64-66]).

From Research to Decision-making

With this confluence, it is not surprising that agencies like the U.S. Department of Defense have made substantial dollar investments in social science, including computational modeling and simulation for understanding human social, behavioral and cultural patterns [67]. National security decision makers, including those in the Department of Defense, can highlight a number of ways in which they would like to use computational social science techniques, including training simulations, characterization of adversary networks and situational awareness. Among these, the ability to forecast is an implicit goal of many projects (see for example discussion on page 25 in [68]). The expectation is that social science-based modeling and simulation tools can be used to forecast future social, political, and cultural trends and events; and that these forecasts will improve decision-making.

Computational modeling and simulation technologies have played an important role in a wide range of human knowledge activities, from academic research to organizational decision-making. The utility of these technologies has been demonstrated over several decades of development and deployment in multiple fields, from weather forecasting to experimental physics to finance. However, it is important to remember that computational modeling and simulation tools are ultimately human artifacts, and like all human artifacts they come with very real limitations. How we recognize and deal with these limitations depends very heavily on the context in which we are using models and simulations. After all, models and simulations have different lifecycles in scientific research contexts than they do in decision-making contexts. Generally speaking, researchers use computational modeling and simulation to support knowledge producing activities:

to refine conceptual models, examine parameter spaces, identify data needs and possible sources to address knowledge gaps. Moreover, models and simulations that are embedded in ongoing cycles of scientific knowledge production benefit from continuous comparisons between empirical data/observations and model outputs, as well as peer review.

Unlike researchers, decision makers often look to modeling and simulation technologies to help refine courses of action that may have very high public consequences. They are frequently dealing with problems characterized by high levels of epistemic uncertainty – i.e., lack of knowledge and data – and are addressing problems for which scientific and expert consensus may be neither mature nor fixed [69]. For decision makers, modeling and simulation technologies may be seen as useful "what if" tools to help them evolve their understanding of a problem space [70, 71]. However, decision makers are probably not focused on improving the model's correctness, or assessing how well it corresponds to a real-world phenomenon of interest. Decision makers tend to be more focused on identifying courses of action and moving forward, and in doing so, they typically face legal, economic, and political motivations and constraints that researchers do not. In the context of the national security community, decision makers may be addressing problems that involve high resource commitments or even human lives.

The contextual difference between research environments and decision-making environments is a critical one that carries significant implications for the design, implementation, and evaluation of computational models and simulations. The decision to employ computational modeling and simulation technologies in high-consequence decision-making implies a responsibility for evaluation: not just of the models themselves, but assessments of how these technologies fit into, shape, and affect outcomes in the real world. Higher consequence decision spaces require proportionally greater attention to assessing the quality of the data, methods, and technologies being

brought to bear on the analysis; as well as the analytic and decision-making processes that rely on these technologies.

In this regard, we briefly highlight three areas of evaluation that we believe require careful attention for computational social science. These include verification and validation (V&V), human-computer interaction, and forecasting as an organizational (not computational) challenge.

Verification and Validation

Verification and validation (V&V) are processes that assess modeling and simulation technologies for internal correctness (verification), and external correspondence to real-world phenomena of interest (validation). There is an enormous body of literature dating back to the 1970s that addresses methods, techniques, tools, and challenges for V&V [72, 73]. Most of this research has been done in fields like computational engineering, artificial intelligence, and operations research. However, in the computational social science community, there is an emerging body of literature addressing the challenges of verifying and validating computational social science models and simulations [14, 74-83]; see also [50, 84].

We will not review the voluminous V&V literature here, except to make two points: firstly, computational social modeling and simulation raises specific V&V issues that are probably unique to the social sciences. Secondly, despite the marked epistemic differences between computational social science and computational physics, engineering, or even operations research, the broader V&V literature does have lessons for organizations investing in predictive computational social science.

Verification and validation in computational physics and engineering is both similar to and divergent from computational social science. For example, in computational science and engineering, determining whether a software tool is accurately solving a set of partial differential equations (verification) is a logically internal process; when large systems of partial differential

equations are typically in play, "correct" in the context of verification means a "mathematically accurate solution." It says nothing about whether or not that solution adequately captures the behavior of a real world phenomenon. As such, verification requires no engagement with the world of observation. Similarly, in the context of agent-based modeling, assessing whether or not an agent-based model accurately executes a conceptual model requires the ability to rigorously assess the mathematics, algorithms, and software engineering of the system. That may require the development of agent-specific verification techniques, but does not require engagement with the external world.

On the other hand, determining whether a partial differential equation is correctly representing a real world phenomenon of interest - that is, performing validation – *does* require engagement with the external world. Correctness in the context of validation must be centered on observations derived from valid sources; i.e., systematic observational data or controlled experiments. Along those lines, assessing whether an agent-based model is built on correct requirements, implementing an appropriate conceptual model, and producing outputs that correspond to the real world, requires comparison with observation.

How we perform a meaningful and objective comparison among the conceptual model, the simulation, and the real world, is a critical challenge in the computational social sciences. For one thing, it is difficult to escape the problem of explanatory/theoretical contingency and plurality in the social sciences, in which cross-disciplinary challenges to explanatory frameworks are common, and demonstrable certainty is rare. Although some might see quantitative modeling as a way of introducing rigor into the social sciences, it is not clear that modeling helps researchers get around this problem. In the realm of the physical sciences, models derive from stable epistemology, rather than vice versa. In the social sciences, there are basic debates about the role of theory as a descriptive, explanatory, or causal framework; and whether or not a nomothetic enterprise is even

possible (i.e., the generation of broadly applicable, generalizable explanatory theories for human behavior). As anthropologist Jessica Glicken Turnley points out, evaluation techniques that rest on a logical positivist philosophy that a) assumes the existence of objective data, and that b) presumes stable relationships between data and theory, are a poor fit for the social sciences, where multiple frameworks can be evoked with equal credibility, depending on one's discipline, to explain similar phenomena [75]. Indeed, evoking computational modeling and simulation to assert epistemological rigor is highly problematic in areas where theoretical consensus is lacking. In particular, confirmation bias is a well-recognized form of cognitive bias in which people subconsciously put greater emphasis on information that is consonant with their reasoning, while simultaneously discounting disconfirming evidence. Insofar as computational models and simulations reify and help us visualize our conceptual models, they can make those models seem more credible than they perhaps are – as critics of computational social modeling projects have pointed out (see for example Andrew Vayda's discussion of Stephen Lansing's work in [85]).

Issues with theory and conceptual validity are intertwined with the problem of data validity, a second challenge for verification and validation in computational social science. In computational physics and engineering, validation depends on two things: identifying a validation referent, or a known point of estimated "truth" for comparison that enables one to evaluate the validation correctness or accuracy of the model vis-à-vis reality; and the ability to generate valid observational data around that referent. In the social sciences, this requirement for known points of truth to act as referents, and the associated need for high-quality empirical validation data, are serious challenges.

In this regard, data will probably be a major, ongoing problem for the verification and validation of computational social models and simulations, since it is impossible to assess the value of a model or simulation without systematic ways to tie the model to observed reality. For one thing, some forms of social knowledge simply resist quantification. At a deeper level, the issue of how

evaluate the "objectivity" of data in the social sciences is a long-standing epistemological debate. This is because social scientists are embedded in the very social matrices they are studying; we cannot speed up or slow down society, or miniaturize it in relation to our senses, to observe the manifold and multilevel dynamics that interest us. As Lucy Resnyansky points out, "Data that are used for understanding the threat of political violence, extremism, instability and conflict are essentially different from what is considered to be data in natural sciences. The former kinds of data have a representational nature and are sociocultural constructs rather than results of objective observation and measuring" ([47]: 42). Lastly, empirical data that are used to develop a model cannot be used to rigorously validate it, which means that validation requires investment in the systematic collection of additional validation quality data. This can be challenging if the phenomenon of interest involves the dissemination of an idea through a large population, or assessing the causes of intergroup violence in a particular region of the world, in which case data collection could easily span many countries and several decades.

This raises a second point: the computational physics and engineering literature that deals with verification and validation is relevant and important for computational social science models and simulations intended for application in real-world decision-making contexts. This literature emphasizes that the main benefit of V&V is not (perhaps counter-intuitively) increased focus on the model, but the contextual issue of how the model will be used, and therefore how the organization and its members identify what decisions they are responsible for making, and negotiate the levels of risk they are willing to accept. This is because verification and validation emphasize whether or not a software application is credible *for an intended area of use*. These discussions force clarification about the decisions, tradeoffs, and risks across stakeholder communities, and what is required for a model to be considered credible and appropriate in relation to a decision. In this regard, we have come to

view verification and validation as a form of sensemaking through which stakeholders in a decision space negotiate the benefits and limitations of a modeling and simulation technology.

Forecasting, Simulation, and Decision-making

A great deal of the literature on computational social science in national security decision-making focuses on challenges of theory, methods and data to support computational modeling and simulation for a range of problems, from training to forecasting. What this focus misses is that forecasting is not a technological problem, and that no model or simulation ever makes a prediction or develops a forecast. Models and simulations generate information. People make predictions and develop forecasts. Whether or not a simulation is actually "predictive" of something is always human judgment, not a technological one, and humans are always in the loop.

In this regard, we call the reader's attention to an extensive body of interdisciplinary scholarship, much of it rooted in economics, business, psychology and management, that focuses on the topic of forecasting and decision-making in organizations (see especially [86-90]). This literature highlights a larger family of forecasting approaches that include quantitative (statistical), qualitative (judgmental), and integrated quantitative-qualitative approaches to developing forecasts. This literature treats modeling and simulation tools technological inputs to forecasting techniques, methods and principles; and emphasizes that tools are only as good as the processes through which they are created and used. In particular, Armstrong identifies eleven different families of forecasting techniques [86-88], and suggests principles for a robust multi-stage forecasting process. Forecasts, he argues, include multiple stages of activity, including formulating a problem, obtaining data, selecting and implementing forecasting methods, evaluating forecasting methods, using forecasts in planning and decision-making, and auditing forecasting procedures to ensure that appropriate principles have been applied [91]; see also [90, 92]. Armstrong's principles point to a kind of "verification and validation" for forecasting beyond the correctness of a model and beg the question

of whether or not a model is actually the best analytic methodology for a particular decision space. Moreover, his work highlights forecasting as an organizational problem, not a technological one; and it is a difficult challenge because planning and decision-making activities tend to be highly distributed within and across stakeholder groups.

No area of research makes this point more thoroughly than weather forecasting, which has been studied extensively by psychologists, decision theorists, and economists for six decades as part of an ongoing effort to assess and increase the political, social, and economic value of weather forecasts. Weather forecasting is unique for several reasons: first, the United States National Weather Service issues many tens of millions of forecasts a year [93]. Second, weather forecasts are highly public, with federal, state and local agencies and individual citizens incorporating weather and climate forecasts into a wide array of daily activities, from purchasing road-clearing equipment to planning weddings. Third, weather forecasters get regular feedback not only on the correctness of their predictions, but on the value of the forecast information they provide. As a result, weather forecasting has been a subject of intense interdisciplinary study for many decades, because weather forecasting is one of the few areas where it is possible not only to evaluate the correctness of a forecast and to suggest improvements, but also to document how forecasts are incorporated into decision-making processes. As Pielke suggests, weather forecasting "provides some lessons about how we think about prediction in general," not just weather forecasting specifically ([93]: 67).

A great deal of this literature is relevant to computational social models and simulations being used for predictive purposes. The weather forecasting literature treats modeling and simulation technologies as only one element of a much larger "process in which forecasters assimilate information from a variety of sources and formulate judgments on the basis of this information" [94]. Moreover, forecasting is not just a problem for meteorologists, but involves a complex ensemble of people, organizations, tools, data sources and activities through which

forecasts are developed, disseminated, acted upon, reviewed and evaluated – what Hooke and Pielke call the "symphony orchestra" of the weather forecasting system [95]. The forecasting orchestra includes three principal activities: forecasting, communication, and incorporation, all of which are working in parallel at any particular point, and each of which can be subjected to rigorous evaluation. Ensuring that this orchestra provides the best public service possible depends on rigorous evaluation of how well each of these activities is performed.

The weather forecasting community not only works to improve the performance of its modeling and simulation tools, but also the skill of the forecasters who develop and disseminate forecasting products. How to evaluate and improve forecasting skill, communicate forecasts, and increase the value of forecasts to decision makers, have been research challenges for meteorologists, psychologists, statisticians, economists and decision theorists since at least the 1960s [94, 96-99]. Forecasting is a process of continuous learning that demands prompt, clear, and unambiguous feedback, in a system that rewards forecasters for accuracy ([100]: 543); forecasters need feedback to identify errors and assess cause [97]. Lacking prompt feedback, intermediate-term feedback can help forecasters get a better sense of how well they are doing; but only when the forecaster's predictions are clearly and precisely recorded, along with the inputs and assumptions or external considerations that went into the forecast. Systematic, regular, comparative evaluation provides more than accountability; it improves forecaster skill.

At the same time, forecasting skill depends not only on the forecaster's cognitive abilities but on "the environment about which forecasts are made, the information system that brings data about the environment to the forecaster, and the cognitive system of the forecaster" ([101]: 579; see also [102]). Thomas Stewart has argued that the forecasting challenge is best understood as an example of the Brunswik lens model, which relates the observed event to the forecast through a lens of "cues" or information items that people use to make the forecast. The quality of a forecast depends

not only on the ecological validity of the cues – that is, how the cues are related to the phenomenon being forecasted and what those cues indicate about the phenomenon – but also the ability of the forecaster to use those cues properly in assessing the event of interest; i.e., whether or not the forecaster is using the right information, and if she is using that information correctly.

As complex as this system is, when all these elements come together properly, weather forecasters are tremendously accurate and reliable in their predictions. However, good forecasting also involves packaging meteorological expert judgment for non-meteorologist consumers. One issue of perennial concern of the forecasting community is the communication of uncertainty in weather forecasts. Forecasting is an inherently uncertain process because of the inexactness of weather science and the many sources of error that can throw off accuracy, including model uncertainty, issues with data, inherent stochasticity, and forecaster judgment. Accordingly, the communication of uncertainty is a major element in whether or not people can use forecasts. In 1971, Murphy and Winkler found that even other scientists had trouble explaining what meteorologists meant by "a 40% chance of rain" [94, 98]. More recent research in human judgment and decision-making indicates that even today, seemingly unambiguous probability statements are prone to misinterpretation: As a simple example, Gerd Gigerenzer and colleagues found that populations in different metropolises interpreted the seemingly unambiguous statement "a 30% chance of rain" in different ways, depending assumptions about the reference class to which the event was oriented [103]. Not surprisingly, the National Oceanic and Atmospheric Administration (NOAA) continues to invest resources in the development of techniques for communicating uncertainty across its stakeholder communities.

Uncertainty is likely to be a major research challenge for forecasts of social phenomena. Research should emphasize methods for quantifying, bounding, aggregating, and propagating uncertainty through both models and the forecasts derived from models. Indeed, a National

Research Council report on dynamic social network analysis identified uncertainty as one of the key under-researched areas in quantitative and computational social science [50]. This research is critical for developing a decision-oriented computational social science, but it is probably not sufficient. If NOAA's experience in this regard is any indication, forecasts of social processes and phenomena will have to deal not only with multiple sources of uncertainty, but also the challenge of representing and communicating uncertainty to consumers with varying levels of skill in interpreting quantitative, graphical, and/or qualitative expressions of uncertainty.

Lastly, it is important to emphasize that forecasting and decision-making are two different activities. That improvements in decision-making do not necessarily depend on improvements in forecasting is illustrated in case studies examining how forecasting failures actually lead to *better* public policy- and decision-making (see for example [104]) All decisions involve uncertainty, both stochastic and epistemic. Putting too much emphasis on forecasting as a means of improving planning can lead decision makers to focus on the correctness of the forecast at the expense of the planning process. Forecasts are helpful as long as they do not divert attention from potentially more robust ways of dealing with uncertainty, such as flexible resource allocation practices or hedging strategies [105].

Users, Transparency, and Responsibility

Verification and validation techniques assess the goodness of a model/simulation from an internal (verification) and external (validation) perspective. In the context of high-consequence decision-making, such as that performed in military and intelligence contexts, there is another dimension that requires assessment. This dimension is the relationship between the model/simulation technology and the person or people using the technology; i.e., the relationship between the human and the computer.

All software projects have various stakeholders, including developers, funders, and end users. In the software engineering community, it is generally understood that getting end users involved in the design and development of the tools they will use is critical if the software is to be usable, useful and relevant to real-world problems. Even so, end users tend to be the silent stakeholder in modeling and simulation projects, because so many begin, progress, and end without much consideration of who will use the software or what they will do with it. We think of this as the "over-the-fence" model of software development. Such over-the-fence software projects are quite common in the national security community.

The over-the-fence model of software development may be particularly poor for computational social modeling and simulation efforts. This is because computational science projects tend to be complicated interdisciplinary efforts that bring together an array of subject matter experts [48]. Very sophisticated models can require deep expertise in a number of areas, from computer hardware to uncertainty in social science data. The process of developing the model is a critical forum for knowledge exchange because model development activities afford developers the chance to learn from each other and to develop shared understandings about the technology under construction [106, 107]. Because of this, we believe that a key challenge for the applied computational social science community is developing relationships with end-users that facilitate transition of modeling and simulation technologies into usable, useful, and adoptable systems that support analytical reasoning.

At a deeper level, this raises the question of how much of this experiential or contextual knowledge is required to effectively use modeling and simulation technology. Because modeling and simulation technologies can embody so many layers of expertise, it can be difficult for end users who are not subject matter experts to understand what the model is doing, or how it performs its functions. Sometimes, this is not an issue because the modeling and simulation technology is not

going to be used outside the domain in which it was developed. It might be a tool that a research or analysis team develops for itself; in this case, the developers are the end users for the technology, and because of that, they understand (hopefully) the model's uses, limitations, and biases. Alternatively, the tool may not be traveling very far outside the domain of its creation. For example, a sociologist might develop an agent-based social network modeling tool, and might post it on her website so that other sociologists trained in these techniques can apply it to their data. In this case, the domain of use is *epistemically adjacent* to the domain of development, so that new users can credibly bring their domain knowledge to bear on the software artifact they are using.

However, when modeling and simulation technologies are going to be transferred across epistemic domains, the question of how and if non-subject matter experts can engage the technology as a tool becomes more problematic. Such *epistemic distance* raises ethical issues for applied computational modeling and simulation projects, since users who do not understand the application space, benefits and/or limitations of a modeling and simulation tool are unlikely to use it well. Along these lines, Fleischmann and Wallace have argued that ethically responsible modeling implies three elements: a commitment to develop models that a) are faithful to reality, b) reflect the values of stakeholders, and c) are maximally transparent so that users and decision makers can employ the model appropriately. This latter property, transparency, is "the capacity of a model to be clearly understood by all stakeholders, especially users of the model" ([108]: 131). Developing processes to deal with epistemic gaps will be an important aspect of tool development and deployment in the national security community. This is an organizational problem, not a technological one, and addressing it requires careful planning and stakeholder negotiations.

Conclusion

As the computational social science community continues to evolve its techniques and approaches, its practitioners may play an important role in shaping our rapidly evolving national security community. In a reflexive way, to the extent that the computational social science community attracts and leverages national security investments, national security topics like terrorism and insurgency warfare are likely to provide major focus areas for the evolution of the field's techniques and specialty areas. In moving computational modeling and simulation technologies out of the realm of research and into the realm of policy and decision-making, we should perhaps consider what is required to develop a realistic, robust understanding of what it means to use models and simulations as decision support tools. We want to reemphasize a point we made earlier: *there is no such thing as a computational prediction.* Computational models and simulations provide outputs, but predictions are a form of human judgments. Computational models and simulations are created by human beings, and like everything we create, our models and simulations reflect (even reify) our state of knowledge at a particular point in time. Focusing our attention on the limitations of models and simulations as tools for human users, and investing resources in assessing what those limitations imply for real-world decision-making, can help us build a stronger understanding of how, where, when, and why computational models and simulations can be useful to people working in fraught, high-consequence decision-making contexts.

WORKS CITED

[1] J. Gleick, *Chaos: Making a New Science*. New York: Penguin Books, 1987.

[2] S. Wolfram, *A New Kind of Science*. Champaign, IL: Wolfram Media, 2002.

[3] N. Eagle, "Mobile Phones as Social Sensors," in *Oxford Handbook of Emergent Technologies in Social Research*, S. N. Hesse-Biber, Ed. New York: Oxford University Press, 2010.

[4] N. Eagle, A. Pentland, and D. Lazer, "Inferring Friendship Network Structure by Using Mobile Phone Data," *Proceedings of the National Academy of Sciences,* vol. 106, pp. 15274-15278, 2009.

[5] N. Eagle and A. S. Pentland, "Reality Mining: Sensing Complex Social Systems," *Personal Ubiquitous Computing,* vol. 10, pp. 255-268, 2004.

[6] D. Watts, "An Experimental Study of Search in Global Social Networks," *Science,* vol. 301, pp. 827-829, 2003.
[7] D. Watts, *Six Degrees: The Science of a Connected Age*. New York: Norton, 2003.
[8] D. Watts, "The 'New' Science of Networks," *Annual Review of Sociology,* vol. 30, pp. 243-270, 2004.
[9] D. Watts and S. Strogatz, "Collective dynamics of small world networks," *Nature,* vol. 393, pp. 409-10, 4 June 1998.
[10] A.-L. Barabási, *Linked: How Everything is Connected to Everything Else and What it Means*. New York: Plume, 2003.
[11] C. Macal and M. J. North, "Tutorial on Agent-Base Modeling and Simulation, Part 2: How to Model with Agents," in *Proceedings of the 2006 Winter Simulation Conference*, L. F. Perrone, F. P. Wieland, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, Eds. Monterey: IEEE, 2006.
[12] D. White, "Networks and Complexity," *Complexity,* vol. 8, pp. 14-15, 2003.
[13] N. Gilbert and P. Terna, "How to Build and use Agent-Based Models in Social Science" 1999. Retrieved June 30 2010 from *web.econ.unito.it/**terna**/deposito/gil_ter.pdf*
[14] L. Yilmaz, "Validation and Verification of Social Processes within Agent-Based Computational Organizational Models," *Computational and Mathematical Organization Theory,* vol. 12, pp. 283-312, 2006.
[15] Defense Science Board, "Report of the Defense Science Board Task Force on Understanding Human Dynamics," Office of the Undersecretary of Defense for Acquisition, Technology and Logistics, Washington, DC 2009.
[16] L. Tesfatsion, "Agent-based Computational Economics: Growing Economies from the Bottom Up." *Artificial Life* 8(1): 55-82.
[17] J. Epstein, *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton, NJ: Princeton University Press, 2006.
[19] T. Koehler, "Putting Social Sciences Together Again: An Introduction to the Volume," in *Dynamics in Human and Primate Societies*, T. Koehler and G. Gumerman, Eds. New York: Oxford University Press, 2000.
[20] P. Ormerod, *The Death of Economics*. New York: Wiley, 1995.
[21] J. M. Epstein, "Agent-Based Computational Models and Generative Social Science," *Complexity*. 4(5): 41-60*,* 1999.
[22] J. D. Farmer and D. Foley, "The Economy Needs Agent-Based Modeling," *Nature,* vol. 460, p. 685, 2009.
[23] J. Epstein, "Modeling to Contain Pandemics," *Nature,* vol. 460, p. 687, 2009.
[24] S. Ressler, "Social Network Analysis as an Approach to Combat Terrorism: Past, Present and Future Research," *Homeland Security Affairs,* vol. 2, 2006.
[25] V. Krebs, "Uncloaking Terrorist Networks," in *First Monday*, vol. 7(4), 2002, retrieved June 30, 2010 from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/941/863
[26] P. R. Keefe, "Can Network Theory Thwart Terrorists?," in *New York Times*. 12 March 2006, retrieved 31 July 2010 from http://www.nytimes.com/2006/03/12/magazine/312wwln_essay.html
[27] J. Bohannon, "Counterterrorism's New Tool: Metanetwork Analysis," *Science,* vol. 325, pp. 409-411, 2009.
[28] JASON, "Rare Events," MITRE Corporation, McLean, VA 2009.
[29] R. Heuer, "Psychology of Intelligence Analysis," Washington, DC: Central Intelligence Agency, 1999.

[30] National Commission on Terrorist Attacks Against the United States, *The 9/11 Commission Report*. Washington, DC: United States Government Printing Office, 2004.
[31] E. Reid and H. Chen, "Mapping the Contemporary Terrorism Research Domain: Research, Publications, and Institutions Analysis," in *IEEE International Conference on Intelligence and Security Informatics* Atlanta, GA: Springer-Verlag, 2005.
[32] B. G. Silverman, G. Bharathy, B. Nye, and R. J. Eidelseon, "Modeling Factions for "Effects Based Operations," Part I: Leaders and followers," *Computational and Mathematical Organization Theory,* vol. 13, pp. 379-406, 2007.
[33] K. Carley, "Destabilizing Terrorist Networks," in *Proceedings of the 8th International Command and Control Research and Technology Symposium*, National Defense War College, Washington DC, 2003.
[34] K. Carley, D. Fridsma, E. Casman, N. Altman, C. J., B. Kaminsky, D. Nave, and A. Yahja, "BioWar: Scalable Multi-Agent Social and Epidemiological Simulation of Bioterrorism Events," in *NAACSOS Conference 2003*, Pittsburgh, PA, 2004.
[35] L. Kuznar, A. Astorino-Courtois, and S. Canna, "Assessing the Perception-to-Intent-to-Action Dynamic," in *Topical Strategic Multi-Layer Assessment (SMA), Multi-Agency/Multi-Disciplinary White Papers in Support of Counter-Terrorism and Counter-WMD* Washington, DC: United States Department of Defense, 2009.
[36] S. Koschade, "A Social Network Analysis of Jemaah Islamiyah: The Applications to Counterterrorism and Intelligence," *Studies in Conflict and Terrorism,* vol. 29, pp. 589-606, 2006.
[37] M. Dombroski, P. Fischbeck, and K. Carley, "Estimating the Shape of Covert Networks," in *Proceedings of the 8th International Command and Control Research and Technology Symposium*, National Defense War College, Washington DC, 2003.
[38] J. C. Bohorquez, S. Gourley, A. R. Dixon, M. Spagat, and N. F. Johnson, "Common Ecology Quantifies Human Insurgency," *Nature,* vol. 462, pp. 911-914, 2009.
[39] M. Lazaroff and D. Snowden, "Anticipatory Models for Counter-Terrorism," in *Emergent Information Technologies and Enabling Policies for Counter-Terrorism*, R. L. Popp and J. Yen, Eds. Hoboken, NJ: John Wiley and Sons, 2006, pp. 51-73.
[40] R. L. Popp and J. Yen, *Emergent Information Technologies and Enabling Policies for Counterterrorism*. Hoboken, NJ: John Wiley and Sons, 2006.
[41] G. Ackerman, A. Battacharjee, M. Klag, and J. Mitchell, "Literature Review of Existing Terrorist Behavior Modeling," Center for Nonproliferation Studies, Monterey Institute of International Studies, Monterey, CA 2002.
[42] H. Goldstein, "Modeling Terrorists: New Simulators Could Help intelligence Analysts Think Like the Enemy," *IEEE Spectrum,* pp. 26-35, 2006.
[43] L. A. Kuznar and J. M. Lutz, "Risk Sensitivity and Terrorism," *Political Studies,* vol. 55, 2007.
[44] E. P. Mackerrow, "Understanding Why: Dissecting Radical Islamic Terrorism with Agent-Based Simulation," *Los Alamos Science,* pp. 184-191, 2003.
[45] S. R. Corman, "Using Activity Focus Networks to Pressure Terrorist Organizations," *Computational and Mathematical Organization Theory,* vol. 12, pp. 35-49, 2006.
[46] R. L. Popp, D. Allen, and C. Cioffi-Revilla, "Utilizing Information and Social Science Technology to Understand and Counter the Twenty-First Century Strategic Threat," in *Emergent Information Technologies and Enabling Policies for Counter-Terrorism*, R. L. Popp and J. Yen, Eds. Hoboken, NJ: IEEE/John Wiley and Sons, 2006.
[47] L. Resnyansky, "The Internet and the Changing Nature of Intelligence," *IEEE Technology and Society Magazine,* vol. Spring 2009, pp. 41-48, 2009.
[48] L. Resnyansky, "Social Modeling as an Interdisciplinary Research Practice," *IEEE Intelligent Systems,* vol. July/August 2008, pp. 20-27, 2008.

[49] J. Epstein, "Modeling to Contain Pandemics " *Nature,* vol. 460, p. 687, 2009.
[50] R. Breiger, K. Carley, and P. Pattison, "Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers," Washington, DC: National Academies Press, 2003.
[51] I. M. McCulloh, M. L. Webb, J. L. Graham, K. Carley, and D. B. Horn, "Change Detection in Social Networks," United States Army Research Institute for Behavioral and Social Sciences, Arlington, VA 2008.
[52] P. V. Fellman and R. Wright, "Modeling Terrorist Networks: Complex Systems at the Mid-Range," in *Joint International Conference on Ethics, Complexity and Organizations* London School of Economics, London, England: http://www.psych.lse.ac.uk/complexity/Conference/FellmanWright.pdf, 2007.
[53] N. Choucri, D. Goldsmith, S. E. Madnick, D. Mistree, J. B. Morrison, and M. Siegel, "Using System Dynamics to Model and Better Understand State Stability," *MIT Sloan Research Paper No 4661-07,* vol. Available at SSRN: http://ssrn.com/abstract=1011230, 2007.
[54] M. Genkin and A. Gutfraind, "How do Terrorist Cells Self-Assemble? Insights from an Agent-Based Model," *Cornell University White Paper,* vol. Available at SSRN: http://ssrn.com/abstract=1031521, 2007.
[55] H. V. D. Parunak and J. J. Odell, "Representing Social Structures in UML," in *Lecture Notes in Computer Science: Agent-Oriented Software Engineering II*: Springer Berlin/Heidelberg, 2002, pp. 1-16.
[56] E. Elliott and L. D. Kiel, "A Complex Systems Approach for Developing Public Policy Towards Terrorism: An Agent-Based Approach," *Chaos, Solitons and Fractals,* vol. 20, pp. 63-68, 2003.
[57] F.-Y. Wang, K. M. Carley, D. Zeng, and W. Mao, "Social Computing: From Social Informatics to Social Intelligence," *IEEE Intelligent Systems,* vol. 22, pp. 79-93, 2007.
[58] L. A. Kuznar and W. Frederick, "Simulating the Effect of Nepotism on Political Risk Taking and Social Unrest," *Computational and Mathematical Organization Science,* vol. 13, pp. 29-37, 2006.
[59] B. G. Silverman, G. Bharathy, B. Nye, and T. Smith, "Modeling Factions for "Effects Based Operations," II: Behavioral Game Theory," *Computational and Mathematical Organization Science,* vol. 14, pp. 120-155, 2008.
[60] K. Carley, "Destabilization of Covert Networks," *Computational Mathematical and Organization Theory,* vol. 12, pp. 51-66, 2006.
[61] M. Tsvetovat and M. Latek, "Dynamics of Agent Organizations: Application to Modeling Irregular Warfare," *Lecture Notes in Computer Science: Multi-Agent Based Simulation IX,* vol. 5269, pp. 60-70, 2009.
[62] N. Memon, J. D. Farley, D. L. Hicks, and T. Rosenorn, *Mathematical Methods in Counterterrorism*. Vienna: Springer, 2009.
[63] I. M. Longini, M. E. Halloran, A. Nizam, Y. Yang, S. Xu, D. S. Burke, D. A. T. Cummings, and J. M. Epstein, "Containing a Large Bioterrorist Smallpox Attack: A Computer Simulation Approach," *International Journal of Infectious Diseases* vol. 11, pp. 98-108, 2007.
[64] S. Weinberger, "Pentagon Turns to 'Softer' Sciences," *Nature,* vol. 464, p. 970, 2010.
[65] N. Gilbert, "Modelers Claim Wars are Predictable," *Nature,* vol. 462, p. 836, 2009.
[66] M. Sageman, *Understanding Terror Networks*. Philadelphia University of Pennsylvania Press, 2004.
[67] "Hearing Charter: Role of the Social and Behavioral Sciences in National Security," in *US House of Representatives Committee on Science and Technology, Subcommittee on Research and Science Education and Committee on Armed Services, Subcommittee on Terrorism, Unconventional Threats, and Capabilities* Washington, DC: United States Congress, 2008, p. 5.

[68] N. K. Hayden, "Dynamic Social Network Analysis: Present Roots and Future Fruits," Advanced Systems and Concepts Office, Defense Threat Reduction Agency, Ft. Belvoir, VA July 2009.
[69] S. Jasanoff, "Contested Boundaries in Policy-Relevant Science," *Social Studies of Science,* vol. 17, pp. 195-230, 1987.
[70] S. Bankes, "Exploratory Modeling for Policy Analysis," *Operations Research,* vol. 41, pp. 435-449, 1993.
[71] S. Bankes, "Reasoning with Complex Models using Compound Computational Experiments and Derived Experimental Contexts," in *CASOS 2002* Pittsburgh, PA: Carnegie Mellon University, 2002.
[72] D. J. Aigner, "A Note on Verification of Computer Simulation Models," *Management Science,* vol. 18, pp. 615-619, July 1972 1972.
[73] R. G. Sargent, "An Expository on Verification and Validation of Simulation Models," in *Procedings of the 1985 Winter Simulation Conference*, 1985, pp. 15-22.
[74] N. K. Hayden, "Verification and Validation of Social Science Simulations: Some Thoughts for Consideration in National Security Applications," Defense Threat Reduction Agency/Advanced Systems and Concepts Office (DTRA/ASCO) 2007.
[75] J. G. Turnley, "Validation Issues in Computational Social Simulation," Galisteo Consulting Group, 2004, p. 8.
[76] R. Axtell, R. Axelrod, J. M. Epstein, and M. D. Cohen, "Aligning Simulation Models: A Case Study and Results," *Computational and Mathematical Organization Theory,* vol. 1, pp. 123-141, 1996.
[77] G. Fagiolo, P. Windrum, and A. Moneta, "Empirical Validation of Agent-Based Models: A Critical Survey," Laboratory of Economics and Management, Sant'Anna School of Advanced Studies, 2006, retrieved July 31 2010 from www.lem.sssup.it/WPLem/files/2006-14.pdf
[78] A. Yahja, "Simulation Validation for Societal Systems" CASOS Paper, Carnegie Mellon University, Pittsburgh, PA 2006, retrieved 31 July 2010 from www.casos.cs.cmu.edu/publications/papers/CMU-ISRI-06-119.pdf.
[79] U. Wilenski and W. Rand, "Making models match: Replicating an agent-based model," *Journal of Artificial Societies and Social Simulation,* vol. 10 2007.
[80] P. Windrum, G. Fagiolo, and A. Moneta, "Empirical Validation of Agent-Based Models: Alternatives and Prospects," *Journal of Artificial Societies and Social Simulation* vol. 10, 2007.
[81] L. McNamara, T. G. Trucano, G. A. Backus, S. A. Mitchell, and A. Slepoy, "R&D For Computational Cognitive and Social Models: Foundations for Model Evaluation through Verification and Validation," Sandia National Laboratories, SAND2008-6453, Albuquerque, NM 2008.
[82] S. Moss, "Alternative Approaches to the Empirical Validation of Agent-Based Models " *JASSS,* vol. 11, 2008.
[83] K. M. Carley, "Validating Computational Models," Carnegie Mellon University 1996.
[84] G. L. Zacharias, J. MacMillan, and S. B. Van Hemel, *Behavioral Modeling and Simulation: From Individuals to Societies*, Washington, DC: National Academies Press, 2007.
[85] A. P. Vayda, *Explaining Human Actions and Environmental Changes*. Lanham, MD: AltaMira Press, 2009.
[86] S. J. Armstrong, "Findings from Evidence-Based Forecasting: Methods for Reducing Forecast Error " *International Journal of Forecasting* vol. 22, pp. 583-598, 2006.
[87] J. S. Armstrong, *Long Range Forecasting: from Crystal Ball to Computer*. New York: John Wiley; available at hops.wharton.upenn.edu/forecast, 1985.

[88] J. S. Armstrong, *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Boston: Kluwer, 2001.
[89] S. A. DiLurgio, *Forecasting Principles and Applications*. New York: McGraw-Hill, 1998.
[90] J. E. Cox and D. G. Loomis, "Diffusion of Forecasting Principles through Books," in *Principles of Forecasting: A Handbook for Researchers and Practitioners*, S. J. Armstrong, Ed. Boston: Kluwer, 2001.
[91] J. S. Armstrong, "Standards and Practices for Forecasting," in *Principles of Forecasting: A Handbook for Researchers and Practitioners*, J. S. Armstrong, Ed. Boston: Kluwer, 2001.
[92] L. J. Tashman and J. Hoover, "Diffusion of Forecasting Principles through Software," in *Principles of Forecasting: A Handbook for Researchers and Practitioners*, J. S. Armstrong, Ed. Boston: Kluwer, 2001, pp. 651-676.
[93] R. A. Pielke, Jr., "The Prediction Hall of Fame," *WeatherZine,* vol. 20, pp. 1-2, 2000.
[94] A. H. Murphy and R. L. Winkler, "Forecasters and Probability Forecasts: The Responses to a Questionnaire," *Bulletin of the American Meteorological Society,* vol. 52, pp. 158-165, 1971.
[95] W. H. Hooke and R. A. Pielke, " Short Term Weather Prediction: An Orchestra in Need of a Conductor," in *Prediction: Science, Decision-making and the Future of Nature*, D. Sarewitz, R. A. Pielke, Jr., and R. Byerly, Jr. , Eds. Washington, DC: Island Press, 2000.
[96] A. H. Murphy and E. S. Epstein, "A Note on Probability Forecasts and "Hedging"," *Journal of Applied Meteorology,* vol. 6, pp. 1002-1004, 1967.
[97] A. H. Murphy and E. S. Epstein, "Verification of Probabilistic Predictions: A Brief Review," *Journal of Applied Meteorology,* vol. 6, pp. 748-755, 1967.
[98] A. H. Murphy and R. L. Winkler, "Forecasters and Probability Forecasts: Some Current Problems," *Bulletin of the American Meteorological Society,* vol. 52, pp. 239-248, 1971.
[99] J. D. McQuigg and R. G. Thompson, "Economic Value of Improved Methods of Translating Weather Information into Operational Terms," *Monthly Weather Review,* vol. 94, pp. 83-87, 1966.
[100] B. Fischhoff, "Learning from Experience: Coping with Hindsight Bias and Ambiguity," in *Principles of Forecasting: A Handbook for Researchers and Practitioners*, J. S. Armstrong, Ed. Boston: Kluwer, 2001, pp. 543-554.
[101] T. R. Stewart and C. M. Lusk, "Seven Components of Judgmental Forecasting Skill: Implications for Research and the Improvement of Forecasts," *Journal of Forecasting,* vol. 13, pp. 579-599, 1994.
[102] T. R. Stewart, W. R. Moninger, J. Grassia, R. H. Brady, and F. H. Merrem, "Analysis of Expert Judgment in a Hail Forecasting Experiment," *Weather and Forecasting,* vol. 4, pp. 24-34, 1989.
[103] G. Gigerenzer, R. Hertwig, E. van den Broek, B. Fasolo, and K. V. Katsikopoulos, ""A 30% Chance of Rain Tomorrow:" How Does the Public Understand Probabilistic Weather Forecasts?," *Risk Analysis,* vol. 25, pp. 623-629, 2005.
[104] J. M. Nigg, "Predicting Earthquakes: Science, Pseudoscience, and Public Policy Paradox," in *Prediction: Science, Decision-making and the Future of Nature*, D. Sarewitz, R. A. Pielke, Jr., and R. Byerly, Jr., Eds. Washington, DC: Island Press, 2000.
[105] J. S. Armstrong, "Introduction," in *Principles of Forecasting: A Handbook for Researchers and Practitioners*, J. S. Armstrong, Ed. Boston: Kluwer, 2001, pp. 1-14.
[106] O. Barreteau, "Our companion modeling approach," *Journal of Artificial Societies and Social Simulation,* vol. 6, 2003.
[107] W. Dare and O. Barreteau, "A Role Playing Game in Irrigated System Negotiation: Between Playing and Reality," *Journal of Artificial Societies and Social Simulation* vol. 6, p. <http://jasss.soc.surrey.ac.uk/6/3/6.html> 2003.

[108] K. R. Fleischmann and W. Wallace, "Ensuring Transparency in Computational Modeling," *Communications of the ACM,* vol. 52, pp. 131-134, 2009.

APPENDIX B: WORKSHOP AGENDA

CONTEXT STATEMENT AND AGENDA
Workshop on Challenges in Computational Social Science
Sponsored by the Defense Threat Reduction Agency and organized by Sandia National Laboratories
La Fonda Hotel, Santa Fe, NM
October 25-28th, 2010

This workshop will assemble an interdisciplinary team of experts to identify and examine major challenges to predictive computational social modeling and simulation for high-consequence national security decision-making. The workshop will produce a multi-authored report in which subject matter experts from the social and physical sciences

a) Examine the state-of-the-art in computational social modeling and simulation in both research and decision-making contexts, and
b) Identify challenges facing the field as it transitions to supporting real-world decision-making in national security, and
c) Identify a core set of principles for identifying and pursuing high-quality applied computational social modeling and simulation projects.

We believe these last two points are critical if computational social modeling and simulation technologies are to be useful and beneficial in organizational decision-making processes.

Laura McNamara and Timothy Trucano have drafted a literature review and report that examines the current state of computational social science in relation to national security decision-making. In this report, we argue that that decision makers investing in computational social modeling and simulation technologies should examine how other fields, including weather forecasting and nuclear weapons certification, make use of computational modeling and simulation in organizational decision-making processes. The paper examines a range of issues in this regard, from the utility and usability of modeling and simulation tools, to verification and validation.

The Defense Threat Reduction Agency's (DTRA) Advanced Systems and Concepts Office (ASCO) is putting this paper through peer review through Summer 2010. Our participants will write a 3000 to 5000-word response to the McNamara-Trucano paper, focusing on a key theme related to their area of expertise. Participants will submit their papers by 15 October 2010, and McNamara and Trucano will compile the papers into a single file and issue this to the group two weeks prior to the workshop. We will also assign each paper a discussant. This discussant is another workshop participant who will be responsible for presenting the paper's ideas to the group, identifying questions the paper raises, and leading a feedback exchange for the author. During the workshop, McNamara and Trucano will lead a roundtable exchange in which participants discuss each paper in turn over a two-day seminar. Once the workshop is over, authors will revise their position papers for publication in a report edited by McNamara and Trucano and published by DTRA.

AGENDA

MONDAY, OCTOBER 25th, 2010 – Participants Arrive in Santa Fe

TUESDAY, OCTOBER 26TH

7:30-10:00	**WORKING CONTINENTAL BREAKFAST, STIHA ROOM** *Introductions* – Laura McNamara *The Background Story, or, "How This Workshop Came to Be"* – Laura McNamara and Jennifer Perry *Our Goals and How We'll Get There* – Laura McNamara *Schedule and Questions* – Laura McNamara, All
10:00-10:15	**BREAK**
10:15-11:00	**Discussion #1:** Rob Albro discusses Mike Vlahos' paper
11:00-12:00	**Discussion #2**: Mike Vlahos discusses Mark Bevir's paper
12:00-1:30	**WORKING LUNCH**, Santa Fe Room Summarize and discuss the morning's themes - McNamara
1:30-2:30	**Discussion #3**: Jessica Tunley discusses Lucy Resnyansky's paper
2:30-3:30	**Discussion #4**: Mark Bevir discusses David Sallach's paper
3:30-3:45	**BREAK with afternoon caffeine and sugar-laden treats**
3:45-4:45	**Discussion #5**: Jeff Johnson discusses Rob Albro's paper
4:45-5:15	**Roundtable:** Major Themes from Today's Discussions
5:15-6:30	**BREAK**
6:30-8:30	**WORKING DINNER, SantaCafe, www.santacafe.com** General Discussion of Computational Social Science in National Security Decision-making

WEDNESDAY, OCTOBER 27th, 2010

7:30-9:45	**WORKING CONTINENTAL BREAKFAST, STIHA ROOM** Overview of today's papers Discussion # 6 David Sallach discusses Francois Hemez' paper
09:45-10:00	**BREAK**
10:00-11:00	**Discussion #7**: Phil Huxtable discusses Jean Scholtz' paper
11:00-12:00	**Discussion #8**: Jean Scholtz discusses Phil Huxtable's paper
12:00-1:00	**WORKING LUNCH**, Stiha Room Summarize and discuss the morning's themes - McNamara
1:00-2:00	**Discussion #9**: Francois Hemez on Jessica Turnley's paper
2:00-3:00	**Discussion #10**: Lucy Resnyansky on Jeff Johnson's paper
3:00-3:15	**BREAK with afternoon caffeine and sugar-laden treats**
3:15-4:30	**Wrap Up Roundtable** Jerry Epstein and Benn Tannenbaum: AAAS Policy Seminar Next Steps: Revising Papers and Submitting for AAAS Policy Seminar

DINNER

We have no dinner plans for the second evening, since we wanted to give you a chance to explore Santa Fe. We can, however, make a group reservation for those of you who want to attend dinner together. Santa Fe has plenty of options!

THURSDAY, OCTOBER 28th, 2010 – Participants Depart Santa Fe

APPENDIX C: WORKSHOP PAPERS

Robert Albro, Mining Sentiments: Computational Modeling and the Interpretive Problem of Culture

Presented at the DTRA/Sandia Workshop on Challenges in Computational Social Science. Santa Fe, New Mexico. October 25-28, 2010.

Abstract: Taking off in the mid-2000s, diverse military, intelligence and security agencies and environments, with a wide variety of priorities, have devoted increasing attention, funding and programming to determine the role of sociocultural knowledge for complex problem-solving across an array of parallel efforts. This includes the developing field of computational sociocultural modeling. The use of computational sociocultural modeling as a problem-solving tool, however, points to definite challenges regarding the concept of culture in particular: how it is identified as a problem area for the purposes of modeling; assumptions regarding the sources for, and methods associated with, collection of relevant cultural data; and the ways it is coded for incorporation into computational modeling architecture. Such challenges also draw attention to the interdisciplinarity of the model-building exercise, with respect to the methods, modes of data collection, and the compatibility of different approaches across the social sciences with respect to culture. Within anthropology – historically responsible for the development of the concept of culture – we can point to ongoing debates about the very definition of culture, its utility as an explanatory paradigm, and its uncertain relationship to human behavior.

However, at present the extent of the challenge posed by the culture concept for the computational modeling community is yet to be fully recognized. In particular, this includes: the question of what is meant by cultural data in the first place, the attribution of meaning to cultural data, and the extent to which cultural data correspond to real-world referents. Many of these concerns converge around the evident disconnect that currently exists between the process of cultural data collection, on the one hand, and the process of model construction, on the other, among potential users. If this disconnect can take many forms, most simply, it resides in the fact that data collectors, analysts, and model builders are typically not the same people. Too often, then, the location, form and identity of data are problematically predetermined by policy-driven and mission-specific user or model building, priorities.

This has direct implications for the meaning of culture. The starting point for what is recognized as data too often derives from prevailing doctrinal definitions, as brought together with the technical requirements of model building and of database management systems, rather than any meaningful distinction, as derived from a given cultural context or community. Going forward, more attention should be given to the ways the conceptual underpinnings of computational sociocultural modeling also significantly determine the *meaning* of data in the first place alongside a robust interdisciplinary appreciation for the challenges posed by the culture concept as more than a set of self-evident variables to be coded: in terms of: the definition of units of data, reliability, comparative fungibility, the relative weighting of cultural and non-cultural inputs, and the basic goals of modeling.

Computational sociocultural modeling (hereafter CSC modeling) is at once an important and relatively recent development within the computational social sciences and a growing footprint of the social sciences in Department of Defense (DoD) problem-solving efforts involving sociocultural knowledge.[9] As has been noted, "When U.S. forces invaded Iraq in 2003, the U.S. military was not particularly concerned about the impact of culture on its operations" (Mansoor 2011: 1). This has changed. The emergence of CSC modeling is at once in step with the now well-publicized DoD turn toward so-called "culture-centric warfare," as represented by the counterinsurgency strategies in Iraq and Afghanistan[10] and by the rising importance of "complex operations" other than war,[11] but also an expression of the shape that turn appears to be taking, going forward. The present discussion is not an evaluation of best practices vis-à-vis how the military should go about incorporating or applying so-called cultural knowledge so much as it examines some implications of the shape of the cultural turn in recent years for the case of CSC modeling.

Throughout this discussion I understand CSC modeling as itself composing a particular interpretive scene, that is, as a form of sense making. Understood in this way, the modeling process is a creative exercise in choice-making, where models are partial and selective representations of a given socio-cultural target domain. As such, we should be skeptical of any claim that CSC modeling offers "transparency" from data to decision-making.[12] Instead of treating models as simple

---

[9] While the computational sciences have been around for some time, the emergence of a "data-driven" computational social science has gained momentum only recently (Laser et al., 2009: 721). This is a story for another time. But, for the military, the enthusiasm for computational social science continues a variety of precedents, such as RAND's mid-20th century emphasis on rational choice theory, applied psychology's development of human factors engineering since WWII, and the influence of complexity theory, as a focus upon understanding dynamic self-organizing systems, across a range of disciplines. Given this, here I am concerned with the ways that, more recently, "culture" has become a subject of attention for computational social science in the context of national security.

[10] In addition to such indicators – representing a military doctrinal shift toward culture – as with the new Counterinsurgency Manual (2007) spearheaded by David Petraeus, a sample of literature documenting this turn includes: McFate (2005), Jager (2007), Selmeski (2007).

[11] Complex operations include such activities as: development, diplomacy, humanitarian interventions, stability operations, nation-building, and other such activities, as these are focused on the winning of "hearts and minds" and on intensive non-kinetic engagements with civilian non-combatants.

[12] On several occasions I have heard advocates for different versions of CSC modeling offer this formulation using the word "transparency" (or a comparable version of it), as an explanation of the value-added of CSC modeling for the

interventions in a forecasting pipeline that intercedes decisively between the real world, that is, the empirical facts of the case or objective data, on the one hand, and high-consequence decisions, on the other, we should recognize the modeling process as a complex interpretive scene creatively generative of new knowledge and as participating in a process of selection. My approach encourages a view of stakeholders in modeling as meaning-makers in their own right. As such, here I examine ways that uncritical uses of CSC modeling pose the interpretive dilemma of the "hermeneutic circle."[13] Since CSC modeling is a part of the interpretive scene of decision-making, it becomes important to ask how modeling is understood by users in their work of interpretation. For interpretive opportunities like texts or models, the problem of the hermeneutic circle refers to a circular inability to move beyond our understanding of a given question in terms other than established or received wisdom. (Heidegger referred to our prior "prejudices.") Here, I consider the extent to which the interpretive point of view generated by a given cultural model is in significant degree already built into the model's purposes and architecture, where it anticipates the "problem" it purports to address. My basic question, therefore, is the extent to which, and how, computational models using cultural knowledge might express the values of their primary stakeholders more than they provide access to hitherto unknown or new insights.[14] One implication of this approach to CSC modeling is the exercise of skepticism regarding whether such models are useful for addressing uncertainty.

Here I am not concerned with the entirety of the diverse developments composing the expanding field of CSC modeling. Instead, I focus narrowly on the circumstances of the culture concept, as it becomes the subject of the work of computation social science. In what follows, I

---

military. The issue this raises is that of exactly what role in the decision-making process is assumed for CSC models: as one among a variety of tools available to decision-makers or as a kind of crystal ball.

[13] Hermeneutics is defined as the study of the methodological principles of interpretation. The problem of the hermeneutic circle has been raised most regularly in debates among phenomenologists, in particular, Martin Heidegger (1971), as a problem that suggests the limits of interpretation.

[14] This position has been significantly influenced by the perspectives on CSC modeling taken by McNamara (2010) and Turnley and Perls (2008) respectively.

address the question of culture as a particular form of knowledge in the context of model construction and application. In doing so, I am specifically concerned with what sort of knowledge is identified as useful "cultural knowledge," as this concept is incorporated into the process of CSC modeling. I treat the modeling process as active work done in order to produce particular cultural understandings but also understandings of culture. Rather than CSC modeling as a relatively benign marriage of cultural analysis with new kinds of computing power and potential access to new sources of cultural data, I explore ways that the meaning of cultural data in the DoD context is itself contingent upon the modeling process. This, in turn, has direct implications for the relevance and potential applicability of the different social sciences to this kind of work, given distinct epistemological investments,[15] and suggests possible limits upon interdisciplinarity as a coherent strategy of knowledge production.

Throughout I compare interpretive frameworks of sociocultural anthropology to that of CSC modeling. Reasons for this are several-fold: 1. my own disciplinary training is in anthropology; 2. historically this discipline has been largely responsible for developing the concept of culture in the social sciences; 3. anthropological expertise – in the form of ethnographic data collection and the interpretation of cultures – has been actively sought by diverse military shops in recent years; 4. the discipline of anthropology also has been proactive in addressing the implications of the military's pursuit of more sophisticated cultural knowledge, in the course of which it has begun to develop a corpus of critique of the military's culture concept. The present argument occupies this space of comparative dialogue between anthropology's and the military's respective accounts of culture. 5. The work of CSC modeling has also been pursued as an exercise in interdisciplinarity. And so, a comparison of the ways the culture concept is deployed across distinct disciplinary commitments –

[15] Epistemology gives attention not to what we know but to how we know it. In this sense, I am concerned with the ways that different social sciences, including anthropology and computational social science, promote different – and sometimes incompatible – ways of knowing through their characteristic methods, tools, and conceptual apparatus.

in this case between the cluster of disciplines composing the cognitive sciences and interpretive or symbolic anthropology – helps us to get a clearer handle on what is at issue in the promotion of, and skepticism about, the work of CSC modeling.

Some disciplinary sources for the "culture" of CSC modeling

One call heard regularly about CSC modeling is that, as a developing field, it is in need of a comprehensive theory of culture, that is, a "coherent conceptual scheme for the study of culture" (Panzarasa and Carley 2005: 1). This call is often accompanied by the complementary need for a "shared framework for data." This need, often voiced, is yet further specified, in terms of the goal of making data interoperable, easily transferrable, and able to be merged, ideally as incorporated into computer-based database management software. As Major General Michael Flynn, currently the top military intelligence officer in Afghanistan, recently put it, "There is a mountain of information. And we aren't doing enough to capture that information…[We need] to share information across all barriers and more quickly."[16] This assumption about information, as variable in its content but as epistemologically equivalent data, at once expresses the felt DoD urgency to control the information battlefield while also underwriting the concern in CSC modeling to establish "common vocabularies" alongside "universally accepted taxonomies" (Numrich and Tolk 2010), including for cultural data. As one DoD program manager who also funds CSC modeling put it, "Misinterpretations easily result from lack of common vocabularies" and "progress requires a common framework and perspective" (Estabrooke 2009). But what is often presented as a logistical priority and needed effort to develop a shared lexicon for an emergent field – the computational

[16] Quote taken from presentation made by Major General Flynn as part of the National Research Council's Workshop on "Unifying Social Frameworks: Sociocultural Data to Accomplish Department of Defense Missions," Washington, D.C., August 17, 2010.

modeling of culture – in fact also has fundamental consequences for what the concept of culture – as a unitary [something] – is taken to be in the modeling context going forward.

The call for information sharing (as a tactical battlefield goal) makes common cause with the need for a shared vocabulary (now as a conceptual problem of the field) and with the complementary need for a common taxonomy (as a shared code book for culture). I am not suggesting that the military should not be finding better ways to share information, using a common vocabulary and taxonomy. But I am suggesting that such an emphasis, in turn, appears to inform the DoD approach to "culture," as equivalent to a unitary body of knowledge and system of classification of a people, held in common and expressed in its own idiom. In short, we can compare this culture concept to the largely defunct if classic anthropological description of culture as a "total way of life."

The computational study of culture assumes that culture is one or another kind of (dynamic and nonlinear) holistic system. If no longer a prevailing assumption among a majority of anthropologists, holism, as a concept, has historically been a basic ingredient of the discipline's "omnibus" definition of culture, as with E. B. Tyler's formulation of a "complex whole" (see Fischer 2007: 2). Yet, the assumption of "culture as a system" (see Casebeer 2005) appears to be widespread in the CSC modeling community.[17] "Systems thinking" is historically well-established DoD common sense informing the relevance of the sciences for military operations.[18] And it is also a technical requirement for any given computational model, so that it can be validated as internally consistent. Systems-based approaches, further, promise the advantages of: a higher level of abstraction, and of the transcending of disciplinary differences of approach, a common language, and relational thinking, along with holism (see Cummings 1980).

---

[17] When referring to a *system* I mean it in the broadest sense to indicate any organized, and cohesive, complex of interactive elements.

[18] For a trenchant discussion of how this unfolded during throughout the second half of the twentieth century between DoD and the so-called "behavioral sciences," consult Robin (2001).

Culture becomes most legible in the work of CSC modeling as a system, where the goal quickly becomes understanding the whole of it – as a central tendency, cultural prototype of some kind, or a set of "core cultural values" – through a working out of the systematic relations among the parts. If this is an agent-based model, these parts are the aggregate of a collection of interacting, boundedly-rational, autonomous, and adaptive decision-makers with "sentiments, opinions and beliefs,"[19] which are built into the model as the agent's underlying motivation. The relations among parts are also often understood by way of analogies with language (grammar), with geography (maps), as well as with computer programming itself (codes). Kathleen Carley, an influential voice among CSC modelers, has sought to describe culture as a "collective cognitive system." that possesses "emergent properties" shaped by underlying "grammars" governing interaction among the parts. In somewhat mystifying fashion and more elaborately, she has also described culture as "networks of minds glued together by cognitive chains of interactive thinking" (see Panzarasa and Carley 2005: 3). We should note an equation of underlying behavioral motivations with an underlying grammar. And such ideas inform her current work, which pursues the realistic modeling of "complex socio-technical systems where people, groups, their ideas, beliefs, and activities co-evolve" (Carley 2009: 13).

As this brief summary suggests, CSC modeling as a computational exercise has its roots in the cognitive branches of the social sciences. For cognitive anthropology, the culture concept corresponds to a mental phenomenon,[20] cognitively organized, and usually represented as a system of rules. Typically the structure of language – instead of real-time social discourse – serves as a paradigm for cultural analysis more generally. And culture is treated as analogous to the linguist's

[19] The phrase – "sentiments, opinions and beliefs" – comes from my notes of a meeting (February 24, 2010) with a MITRE project leader for multiple DoD projects, all of which share a focus on how best to transition socio-cultural understanding by way of modeling to the U. S. military. My interest in noting this phrase is to draw attention to the regular grouping of these terms in the work of CSC modeling, where cultural meanings come to be equated with opinions. I develop some further implications of this below.

[20] In the terms of disciplinary history, the equation of culture with cognition has a well-established genealogy traceable to such foundational ideas as Durkheim's "collective representations."

grammar (Foley 1997: 108), similar to Chomsky's "computational system," as representing the grammatical set of possible combinations of signs. As classically represented by componential analysis, characteristic work generative of the cognitivist culture concept includes the collection of sets of words in a given native language assumed to denote distinct categories in specific semantic domain, that is, "folk classification."[21]

Cognitivist starting points such as these also make evident sense to DoD-type behavioral and cognitive science, tasked, in the words of Major General Flynn, with explaining, "What makes people tick?" But such holistic cognitivist and computational approaches, however, are in problematic tension with a more contemporary account of culture among interpretive and symbolic anthropologists, as public (rather than a mental state), historically constituted, open-ended, multiply interpreted, unevenly distributed, and regularly contested.[22] Concern for the culture-as-system-as-a-whole, in other words, risks the gross misrecognition of concertedly non-holistic cultural realities, including the ways meanings circulate through societies, as multi-vocal (that is, as subject to multiple interpretations), and as publicly and dialogically co-constructed. Rather than a political critique, these different assumptions about what culture is and how it works underwrite contemporary anthropological and computational social scientific projects respectively, and help to explain reservations among anthropologists about the leveraging of culture for DoD problem-solving. I will return to the implications of this later.

Doctrinal sources for CSC "culture"

---

[21] Given that our subject is computational modeling, we should note, as well, that such work in the cognitive social sciences historically has actively traded on the metaphor of the human mind as like a computer, that is, as an "opportunistic information processor" (Foley 1997: 115).

[22] This description is at best an abbreviation of a much more protracted, and ongoing, disenchantment with the culture concept within U. S. anthropology in particular, which has taken various forms, and which has been ongoing since the mid-1980s. For further details about anthropology's changing relationship to the culture concept, and discussion of its diminished utility as a unifying disciplinary term, see: Abu-Lughod 1991, Fischer 2007, Scott 2003.

Examples from Carley's work, such as BioWar, a city-scale model for exploring how epidemics behave, are comparable to other agent- or network-based cultural models, such as NonKin Village, an agent-based model developed by Barry Silverman. NonKin Village is a SimCity-like training game that allows users to explore factional scenarios in a "foreign culture" (e. g. Baghdad), while deploying both COIN and SSTR ("Support for Stability, Security, Transition and Reconstruction") "doctrine strategies" (see Silverman et al. 2010).[23] NonKin also functions as a "human terrain data framework," having linked up its human terrain village data with the Marine Corps MarineLink database (Pietrocola 2009: 18). This raises a further issue about the ways CSC models are currently designed and populated. At least some of DoD's CSC programs are now proposing to provide, as one program director put it, "validated models to support human terrain understanding." Such a statement does not imply direct collaboration with the Human Terrain System (HTS) program per se, but goes beyond HTS representing the doctrinal emergence in DoD of a more generic "human terrain analysis" and "human terrain mapping" (see Albro 2010a; Marr et al. 2008), imagined to combine ethnographic-type data with geospatial data and qualitative with highly technological methods of data collection.

This leads to several further implications: 1) DoD's CSC modeling program's are actively selling themselves as a capacity for HTS-like socio-cultural cells to be hosted by the United States Africa Command and other regional combatant commands (e. g. ONR 2010). 2) It has become increasingly routine for CSC models to promote their relevance in terms of the analysis of human terrain, as with a model recently demonstrated to me to explore distributions of violent language on jihadi weblogs and elsewhere, which describes itself as "mapping the sociocultural terrain using

[23] Silverman's NonKin Village model has recently been described in the following terms: "The program, which loosely resembles the game SimCity, is part of a US government effort to develop sophisticated computer models of real Afghan villages – complete with virtual people based on actual inhabitants – in an attempt to predict their reaction to US raids and humanitarian aid" (Stockman 2010).

social and news media."[24] In a comparable manner, Kathleen Carley's CASOS modeling team emphasizes their role as a reach-back cell, intending to feed extracted socio-cultural information back "to field researchers to facilitate rapid assessment of the human terrain" (Carley 2009: 13). Likewise, as researchers from MIT's Lincoln Laboratory have suggested, CSC modeling makes an excellent tool for "human terrain preparation" in order to combat "clash of civilizations"-derived negative attitudes about "the West."[25]

3) But of greater importance is the fact that military doctrine, as with the new doctrinal terms human or "cultural terrain," directly influence the form and function of a given model. During the conversation with a computational modeler following a CSC model demonstration of an "information operations scenario" about how best to "foster tribal support," he underscored the importance of building a model that offers a "holistic view" but that also "fits with the doctrinal constructs of the user community." In fact, for the model in question, the cultural knowledge, how cultural data is collected, and how it is expressed by the model had to conform exactly both to PSYOP and to JIPOE ("Joint Intelligence Preparation of the Operational Environment") doctrine or, in his words, "it won't get used." Whether COIN, PSYOP, SSTR, or JIPOE doctrine, the critical point is that doctrine is used as a starting point for the modeling code, or algorithm, underwriting the computational work of the model itself. The military's definition of culture, in other words, as defined by its doctrinal parameters, provides the grammar for whatever sense-making activities the model is responsible for. But the military's culture doctrine is at best several times removed from any particular cultural reality with which a given model might be concerned. And doctrinal sources for culture have their own conceptual biases designed to encourage familiarity and legibility among the specific user community of military personnel rather than to describe cultures in their own terms.

---

[24] Chen, Hsinchun, "From Dark Web to Geopolitical Web: Collection and Analysis," slide 42.
[25] Lemnios, Zachary J. and Victor Zue, "Research Challenges for the Next Decade," slides 2-3.

An emerging anthropological critique of the military's culture concept has begun to point to some of its anachronistic limits and epistemological biases, while emphasizing notable differences between it and current anthropological engagements with culture (see Davis 2010; Gusterson 2010; Price 2010). For CSC modeling, this can be a concern for ways that the design, the internal structure or architecture, of a model might incorporate diverse "unwarranted assumptions" (Zacharias et al. 2008) regarding the cultural domain. What we appear to have here is a case of CSC models running up against the problem of the hermeneutic circle. The DoD is pursuing the generation of knowledge for decision-making about culture significantly based on self-referential "unwarranted assumptions" built into its own current doctrinal framework for culture, which tell us more about the military and military culture than about other cultures. Fundamentally, this might be a case – in the words of Hugh Gusterson (2010: 279) – of seeking "algorithmic solutions to hermeneutic problems." What this comment dramatizes are different methodological and conceptual commitments about the sources and meaning of cultural knowledge. In the military mode, "cultures" can be modeled in terms of "terrain." In the anthropological mode, cultures are open-ended problems of the negotiation of multiple meanings. These differences are not trivial, and have real consequences for an interpretation of the cultural realities of a given community, country or region.

Particularly for the Marine Corps, successfully navigating the "cultural terrain" has become a critical ingredient for the success of COIN. As a concept cultural terrain was formally leveraged into the Marine Corps beginning in 2004, as the result of work by its Cultural Awareness Working Group among others, which explained that "culture is simply another element of terrain." The Working Group suggested the term be introduced since "terrain" is already a well-established concept "familiar to all military personnel," and so, "cultural terrain is a term that can be used to ease the incorporation of cultural awareness into training, planning, and operations."[26] As is repeatedly

[26] Connable, Ben and Art Speyer, "Cultural Awareness for Military Operations, slide 7.

emphasized when talking about potential roles of CSC models for DoD missions, "Only a concerted effort at defining mission-oriented needs can provide the appropriate framework for modelers" (Numrich and Tolk 2010: 5). At the same time, however, we should think about the ways that attention to the needs of the particular community of users can also begin to drive the very form of the modeling exercise to the detriment of that which is ostensibly the subject of a given model.

One aspect of this problem is the question of "mapping." As Paula Holmes-Eber emphasizes, "Marine learning emphasizes visual representation of information through maps, graphs, and power points." She notes, "Cultural information is often referred to using military geographic terminology" (2007: 5). Modelers have also noted this. As one model-builder explained in reference to the inclusion of geospatial data in his model, "If you can't relate it to a map, and understand how things are visualized, it doesn't work. So, we had to put 'geospatial' in there." And when training prospective users, the most frequent request is, "Let's see it on a map." The need for mapping, in short has less to do with any engagement with previously unknown cultural knowledge so much as it holds a "mirror" (see Rorty 1979) to the ways that military users are encouraged to think about culture in the first place. And the user requirement of visualization has consequences for the technical incorporation of culture into CSC models.[27]

The concept of cultural terrain signifies more than just the introduction of a new lexicon. It incorporates the culture concept into the Marine's knowledge space in ways significantly determining what culture is for. "Culture" has now been incorporated into terrain analysis and defined as "a feature of the terrain that has been constructed by man." As is explained, "Included are such items as roads, buildings, and canals; boundary lines; and, in a broad sense, all names and legends on a

[27] Commenting on the relationship between counterinsurgency, culture and visualization, Mirzoeff (2009: 6) observes, "Culture is itself understood in this contradictory fashion as a totalizing system, governing all forms of actions and ideas in an oscillation between Victorian anthropology and the first-person-shooter video game."

map."[28] It is hard to miss that here culture has become a cartographic asset, with attention directed to culture as a feature of the landscape and as a spatial arrangement to be surveyed. Culture is here significant as it is integrated into a topographic "complex whole." To reliably read the cultural terrain, any behaviors need to stay fixed. People become a spatial array of points and cultural meanings are inscribed in locatable physical places and structures, through which people move and to which they orient themselves. U.S. Marines in theater are supported with a variety of map-like tools to help them read the cultural terrain, beginning with cultural smart cards.[29]

While individual Marines are unlikely to treat culture in such a relentlessly cartographic way, mapping nevertheless does not simply represent the world. It actively produces it. As a feature of the terrain, culture is now particularly available to geospatial technologies dedicated to mapping three-dimensional object space, terrain visualization, GIS integration, and other interoperable data management tools. This makes sense, too, if you want to leverage "every soldier as a sensor" to better enable the availability of the human or cultural terrain knowledge-base across the battle space. Human Terrain Teams, often supporting Marine units, are now being trained to use the MAP-HT/TIGR Tactical Ground Reporting System while on patrol, a map-centric application helping to collect cultural facts more efficiently. Meanwhile, CSC modeling tools such as AutoMap – a text-mining software – automatically classifies information into a spatio-temporal map grid of "social, knowledge, belief, resource, and task networks."

The cultural terrain, as doctrine and as model architecture, is hard to distinguish from the goal of the information control of the battle space. It makes perfect sense for Marines that the efficacy of their culture doctrine is to better equip them to read their environment. As a terrain feature, culture is also subject to technical enhancement to provide a more comprehensive

[28] Joint Publication 1-02, p. 119.
[29] Parts of this discussion of the development of culture doctrine for the Marine Corps (and for the military) can also be found in Albro (2010b).

visualization of it. But conceptions of cultural terrain are unlikely to be relevant frames for cultural meaning among others, particularly the civilian counterparts engaged by U.S. forces and who are the subjects of CSC modeling exercises. With such arrangements, we learn more about modelers and Marines than about anyone else, and we become insulated from the interactive basis of cultural interpretation that is the perpetual context of ethnographic work.[30]

Data extraction and implications for ethnography

CSC modeling programs appear to be ambivalent about the question of data. On the one hand, the concern for data is recognized as a basic problem for future success of computational modeling in DoD, as well as one primary purpose of modeling, which is to provide one reliable means for decision-makers to move "from data to decisions." On the other hand, as I have heard often, these same programs actively distance themselves from the work of data collection. As one program director stated the matter, "We are not a data program!" Indeed, as Numrich and Tolk (2010: 3) note regarding the CSC community, "Nobody wants to be responsible for data."[31] But, given priorities of counterinsurgency and of knowing the terrain, there has been a push to populate CSC models with more grounded cultural details, specifically, ethnographic data.

Modelers, in their turn, have expressed concern that "unstructured narratives" – as these represent the universe of qualitative data collection – are unsystematic and so resist incorporation into modeling architectures. They are usually "not useful."[32] As Turnley and Perls (2008) have pointed out, computational models require data that can be manipulated quantitatively or

[30] As Laura McNamara (personal communication) has suggested, the culture-as-terrain metaphor also has the effect of effacing the ways that cultural understandings are also historically produced. As such, modelers tend not to "acknowledge history" as part of model-building. .

[31] These comments, if frequent, are also part of a larger story and context within DoD and the intelligence community historically, where "collectors" and "analysts" are considered distinct non-overlapping roles. Of course matters are different for anthropologists, where these roles coincide in the same ethnographer.

[32] These quotes are taken from my notes on conversations with different DoD-funded computational modelers during the past year.

algorithmically, while sociocultural knowledge typically takes narrative form. In this sense, computational models actively exert a push away from context-dependent or domain-specific data of all varieties. For example, semi-regular statements of the goal of "relative data completeness" operate epistemologically with an empirical conceit that meanings are somehow contained in vehicles, for extraction. And yet, as the point is made, "Methods are needed to transform ethnographic findings about society's self-perceptions and decision-making skills into frameworks that can inform behavior models."[33] Such a call is commensurate with the goal of a "rich" representation of data composing the virtual modeling environment. But, we need to think further about what sorts of challenges ethnography fundamentally poses for CSC modeling efforts.

Suggesting ways to fix cultural intelligence gathering in Afghanistan, Flynn (2010) has emphasized a preference for detailed and more ethnographic-like "district narrative assessments."[34] And as already discussed for the modeling efforts of Carley, Silverman, and others, one project has been comprehensive incorporation of qualitative human terrain data from the field into models. Modelers have, therefore, begun to look to automation to speed up the extraction of "data" from "texts," including the development of new semi-automated "rapid ethnographic retrieval" systems.[35] But such technologies come with a price. We should note that model requirements are now driving qualitative data collection. In similar fashion, human terrain teams have been switching from "open-ended reports to more rigid questionnaires that can easily be uploaded into a database" (Stockman

[33] This quote was taken from a 2008 Broad Agency Announcement circulated to attract proposals for the Office of Naval Research's "Human Social Culture Behavior Modeling Program."
[34] We might note the evident similarities between Flynn's proposal, in the context of an ongoing occupation, and colonial era British efforts to do the same, where British social anthropologists were in the regular business of collecting data to be used in "handbooks of law and customs," used in turn by British colonial administrators as blueprints for dealing with native populations in sub-Saharan Africa, India, and elsewhere. This is the source of the original charge against anthropology as a "handmaiden to colonialism."
[35] Quoted in Carley (2009: 13).

2010).[36] Here compatibility and inter-operability are driving the relationship between data collection, model building, and analysis, rather than the process of inquiry itself.

Hard-to-classify "field notes" now quickly take the form of more standardized "field reports," which can ideally be quickly scored with a commonly used "code book" of some sort like the popular ASCOPE (Area, Structures, Capabilities, Organizations, Peoples, and Events) system for the classification of field data.[37] Relatively thin more easily extractable data sources are now given priority, such as available measures of public opinion, news articles, press releases, national polling data, or information posted on websites. "Ethnographic field notes" are also discussed by modelers, but in the words of Kathleen Carley (2009: 13), in ways "similar to the type of factors currently extracted from the human relations area files." This can easily become decontextualized cultural "content knowledge," to be counted, aggregated, grouped, and finally archived. With this concept of ethnography, models are given a task we might describe as the generation of significant information about a patchwork world of data points – like a series of cultural boxes to check off – representing quantifiable variables of often pre-assigned cultural relations. In most cases, qualitative data is shoe horned into the DoD epistemology for information, as at once: extractable, mergeable, subject to a common lexicon, and reliably managed by interoperable databases. A concern for requirements for use of cultural data here appear to be prioritized over and above the significance attributed to culture by the cultural subjects themselves.

Ethnographic data is being leveraged in DoD, in keeping with overriding goals for information about populations, in ways that increasingly emphasize "extraction" to the detriment of what we might call the "dialogic ground" (Tedlock and Mannheim 1995) composing the scene of the work of ethnography, as it is understood by contemporary anthropologists, where cultural meanings

[36] In making this observation, I am setting aside some evident problems associated with the HTS program, in terms of recruitment, training, ethnographic practice and ethics. This does not, however, diminish those problems.
[37] ASCOPE is a qualitative data collection technique available to civil affairs personnel and in which human terrain team members are currently trained.

are outcomes of reciprocal negotiations with counterparts in the field. The key distinction here is between pre-encoded and habitual representations available for collection and removal and dialogic practice as contextual meaning-making. Another way of making this distinction is that between identification of "immobilized variables" (where "their culture" is composed of "x" features) (see Fischer 2007: 1) and a more inquiry-driven process attuned to the possibilities of unexpected or new kinds of cultural understanding outside of any given meaning horizon..

The DoD need to pursue specific kinds of population-centric information, identified as a prior goal, therefore, drives an increasingly less open-ended more scripted ethnographic engagement that generates knowledge already determined to be important, in standard formats, rather than the negotiated serendipities of "thick description"[38]-style ethnography that is particularly suited to the study of unpredictable outcomes, for which typically where you start is not where you end up. We might be face to face with what Robert Rubinstein has called the "fallacy of detachable cultural descriptions."[39] In the absence of counterparts' voices engaged with our own, or any context other than our own requirements and priorities, we tend to highlight differences between *us* and *them* by objectivizing other cultures as distinctive and ourselves as normal. In this way, CSC modeling can become a technology for the production of cultural difference as objective fact in ways that distance us from an appreciation of others' interpretations of their own cultural realities or of us.

Opinion mining, new social media and sentiment classification

As a computational modeling professional noted to me, these days "everyone is pushing for extraction." Part of the reason for this is the difficulty of populating CSC models with hard-to-collect "thick description" ethnographic data, which seeks not only to describe cultural behavior, but

[38] The term – thick description – is borrowed from Clifford Geertz's (1973) famous essay, "Thick Description: Toward an Interpretive Theory of Culture," which in turn borrowed from Gilbert Ryle, to contrast ethnographic descriptions "thick" and "thin."
[39] Robert Rubenstein: personal communication.

to provide its meaningful context. Therefore, modelers have looked for alternatives. The most promising of these, from the modeling point of view, is largely open source online digital information. Of particular interest is Web2.0-type social media content, as generated by different "target groups." Using sophisticated web crawling and spidering software suites, modelers can collect terabytes of data from a wide range of platforms, from web sites, blogs, YouTube, Second Life, and HTML pages of all sorts, and without having to interact with people. The extracted and compiled data comes from simple Word documents, PDF files, PowerPoint presentations, image, video and audio files, among other kinds of content. A variety of further analytical activities can be performed on this massive data source. But here I address just one such activity, increasingly a part of CSC modeling collection and analysis, referred to as "sentiment extraction."

The agents that compose CSC models are regularly described as having been given realistic "sentiments, opinions, and beliefs." There are wider and narrower versions of this claim, ranging from "preferences," "sentiments," or "choices" to the inclusion of the understanding of "motivations," "genre," "style," value systems," and even including "sub-cultures."[40] One CSC modeling program has described its priority as the "extraction and attribution of sentiments to peoples and cultures." And given that our consideration is the incorporation of cultural knowledge into CSC modeling, here I give attention to the ways that *sentiment* is used often largely interchangeably with *cultural belief* in the work of sentiment extraction and classification. While, in fact, the two concepts are notably different, regular slippage between these several synonyms amounts to an additional way that culture is smuggled into modeling's interpretive scene. However, in so doing, the computational process grossly misrecognizes the significance of culture, including the holistic "total way of life" version of culture that otherwise has such traction in this work.

[40] For discussion of "sub-cultures," see Lemnios and Zue, op cit., slide 10. The rest of these terms have been gathered from across multiple primary and secondary sources, as terms used in overlapping ways both with "sentiment" and with "cultural meanings."

The computational work of sentiment extraction compiles largely textual data. So-called sentiment analysis is understood by modelers as fundamentally a text classification problem. This classification work can be performed at the level of an entire document, a sentence, or sentence clause. As Liu (2010: 267) notes, "Natural language documents are regarded as unstructured data," which is subject to traditional data management tools "to be applied to slice, dice, and visualize the results in many ways." Sentiment, in this case, refers to usually subjective opinions. Typically, such "opinionated text" is understood to express either a positive or negative orientation. And the work of sentiment extraction and classification boils down to the identification of the orientation of opinionated texts (or sentences or clauses), understood to be relationally contrastive (e. g. favor/disfavor, like/dislike, accept/reject, etc.).[41] Modelers seek to identify orientation indicators using a "bag of words" approach that identifies grammatical features or lexemes like adjectives and adverbs. Once identified, these lexemes are assigned a value. The arrangement of these relationships can then be mapped and represented as a tree, hierarchy, or taxonomy.

Such a sentiment analysis mining the newly massively available user-generated Web content first became relevant in market research, as models collected and interpreted information indicating consumer attitudes about particular products. It has only recently been brought over for use in the DoD context. We should, however, note that sentiment classification, as a computational means of coming to terms with culture, has its origins in market research rather than any thinking about culture per se. What appears to be the case is that *opinion* – understood as subjective expressions, also often called "feelings and beliefs," of pro/con with respect to something else – is being treated as synonymous with cultural *meanings*.

[41] While I offer only a basic gloss of sentiment extraction and classification here, as part of the full suite of tools used in CSC modeling, this process is subject to considerably greater sophistication, as texts, or sentences, can be further separated into component parts, with each possessing multiple attributes. Adding to this complexity is the fact that sets of sentiment indictors can also be fuzzy. For further details see Andreevskaia and Bergler 2006; Liu 2010; Yi and Niblack 2005.

It is important, furthermore, to recognize that sentiment classification requires a bootstrapping approach, and often uses online dictionaries like WordNet, to first grow a lexical or semantic word set composed of antonyms, synonyms, or other relational terms, which the model then utilizes to evaluate and organize opinion polarity in unstructured online natural language texts. Currently CSC modelers promise ever greater realism in such areas as the understanding of complex religious motivation, economic incentives, and the significance of tribal affiliation. Yet, sentiment classifiers admit that at present it remains particularly challenging effectively to relate opinion words to domain specific contexts. One way to understand this is that – hermeneutic circle-like – such sentiment classification relies upon already established classificatory sets of relations, independent of the model's work of interpretation, and becomes difficult or ineffective when seeking to make sense of diverse vernacular worlds. Furthermore, cultural meanings associated with religious motivation, say, can be understood as context-specific propositions about the cosmos. But, we cannot in turn hope to understand such meanings in their own terms by requiring that they exhibit a pro/con attitude with respect to something else, as if such beliefs were analogous to consumer behavior.

The modeling work of sentiment extraction and classification is familiar, insofar as it rehearses a developed strategy of componential analysis and similar efforts designed to establish the systematicity of semantic contrasts or of binary distinctions (Foley 1997: 108). And, as with componential analysis automated componential sentiment extraction also treats culture, as organized sets of lexemes, as a *code* to be cracked. We should note the regular call among CSC modelers to "create a code book" for non-quantitative data in particular, which includes a "list of all the desired variables and a description of how the variable is to be interpreted" (Carley 2009: 13). Key indicators of "sentiments, opinions, and beliefs," are assumed to infer more fundamental or hierarchically causative "factors that influence their behavior" (Numrich and Tolk 2010: 2), treated as "preconditions."

But as linguistic anthropologists Mannheim and Becker (1995: 238) observe in this regard, the notion that a "code preconstitutes language" means that "utterances and meanings are so to speak always already written." They note that the code metaphor is both "subversive of a dialogic image of language" (1995: 237) at the same time as it leads to what they call the "troping" of other cultures. That is, it makes other cultures appear to be exotic when compared to our own. Computational approaches to sentiment extraction also reveal a basic difference of orientation with respect to where *meaning* resides: on the one hand, a cognitive/computational code that is assumed to underwrite and to determine the systemic properties of cultures; on the other, ethnographic approaches to the public, expressive and active meaning-making of people viewed as "creative cultural producers" (Fischer 2007: 38), characteristic of contemporary interpretive anthropology, where in the spirit of Bakhtin, meaning-making involves the active reframing of the past, or reworking of histories, into present relationships. This difference between habitual motivations and creative expressions is not casual but epistemologically fundamental, and it locates very different and largely incompatible projects for culture.

* * *

The computational modeling of culture potentially offers increasing sophistication and interpretive power, as DoD seeks to make sense of "complex operations." At present just how CSC modeling is best incorporated into the broader interpretive scene for high consequence decision-makers is yet to be sharply defined. Part of this boils down to the technical challenges posed by highly sophistical computational models, which means that laypersons can too easily use them as like "terministic screens"[42] – grids of intelligibility or legibility through which information is passed – in

[42] I am borrowing this term from the literary critic Kenneth Burke.

the work of sense-making, even as they have little detailed understanding of them. At the same time, CSC models can pose challenges to an often assumed social scientific interdisciplinarity, since epistemological commitments of distinct disciplines concerned with culture are not necessarily consistent with those of computational cultural knowledge production, as I have explored by way of its awkward relationship to the semiotic work of contemporary interpretive anthropology. For the case of contemporary anthropology, this includes: an emphasis upon the uneven distributions and multiple interpretations of meanings (instead of holism), dialogism (rather than extraction) and emphasis on creative expression (rather than habitual codes). But, when we consider: the doctrinal sources for cultural knowledge in CSC modeling construction, its attention to user priorities rather than to those of key counterparts (and their accompanying culture), its preference for context-stripping thin and box checking ethnographic data, as well as its use of code books with pre-established values, more fundamentally, a real question remain about the extent to which CSC modeling in fact generates fundamentally new interpretations of anything or whether it is an illustration of a technical engineering-type DoD social science unwittingly caught up in a frustratingly circular and self-referentially contained practice.

REFERENCES

Abu-Lughod, Lila
1991 Writing against Culture. In Recapturing Anthropology: Working in the Present. Pp. 137-154. Richard Fox, ed. Santa Fe, NM: School of American Research.

Albro, Robert
2010a Anthropology and the Military: AFRICOM, Culture, and the Future of Human Terrain Analysis. Anthropology Today 26 (1): 22-24.

2010b Writing Culture Doctrine: Public Anthropology, Military Policy, and World Making. Perspectives on Politics 8 (4): 1087-1093.

Andreevskaia, Alina and Sabine Bergler
2006 Mining WordNet for a Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses. Proceedings of the European Chapter of the Association for Computational Linguistics.

Carley, Kathleen M.
2009 A Dynamic Socio-Cultural Network Lens. HSCB Newsletter No. 2, Summer: 12-13.

Casebeer, William D.
2005 The Importance of Treating Culture as a System: Lessons on Counter-insurgency Strategy from the British Iraqi Mandate. Strategic Insights 4 (10).

Chen, Hsinchun
2010 From Dark Web to Geopolitical Web: Collection and Analysis. PowerPoint briefing. Author's files.

Connable, Ben and Art Speyer
2004 Cultural Awareness for Military Operations. PowerPoint briefing. Author's files.

Cummings, Thomas G.
1980 Systems Theory for Organization Development. John Wiley and Sons, Ltd.

Davis, Rochelle
2010 Culture as a Weapon System. Middle East Report 255 (http://www.merip.org/mer/mer255/davis.html), accessed July 31, 2010.

Estabrooke, Ivy
2009 Directions for Human, Social, Cultural and Behavioral Sciences at the Office of Naval Research. PowerPoint briefing, August 20. Author's files.

Fischer, Michael M. J.
2007 Culture and Cultural Analysis as Experimental Systems. Cultural Anthropology 22 (1): 1-64.

Flynn, Michael T.

2010 Fixing Intel: A Blueprint for Making Intelligence Relevant in Afghanistan. Washington, D.C: Center for New American Security.

Foley, William A.
Anthropological Linguistics: An Introduction. Oxford, UK: Blackwell.

Geertz, Clifford
1973 Thick Description: Toward and Interpretive Theory of Culture. In The Interpretation of Cultures. Pp. 3-30. New York: Basic Books.

Gusterson, Hugh
2010 The Cultural Turn in the War on Terror. In Anthropology and Global Counterinsurgency. J. Kelly and B. Jauregui, eds. Pp. 279-295. Chicago: University of Chicago Press.

Heidegger, Martin
1971 The Origin of the Work of Art. In Poetry, Language, Thought. Albert Hofstadter, trans. New York, NY: Harper Collins.

Holmes-Eber, Paula
2007 Mapping Cultures and Cultural Maps: Representing and Teaching Culture in the Marine Corps. Unpublished paper presented in the conference "Pedagogy for the Long War: Teaching Irregular Warfare." Quantico, VA. October 29-31.

Jager, Sheila Miyoshi
2007 On the Uses of Cultural Knowledge. U.S. Army War College (Carlisle, PA): Strategic Studies Institute Report.

Laser, David et al.
Computational Social Science. Science 323: 721-723.

Liu, Bing
2010 Sentiment Analysis and Subjectivity. In the Handbook of Natural Language Processing. 2nd edition. N. Indurkhya and F. J. Damerau, eds. Pp. 627-666. Chapman and Hall/CRC.

Mansoor, Peter
2011 The Softer Side of War: Exploring the Influence of Culture on Military Doctrine. Foreign Affairs, January/February.

McFate, Montgomery
2005 An Organizational Solution for DoD's Cultural Knowledge Needs. Military Review, July-August: 18-21.

Numrich , S. K. and Andreas Tolk
2010 Challenges for Human, Social, Cultural, and Behavioral Modeling. SCS M&S Magazine 1 (1): 1-9.

Marr, Jack et al.

2008 Human Terrain Mapping: A Critical First Step in Winning the COIN Fight. Military Review, March-April: 18-24.

McNamara, Laura
2010 Why Models Don't Forecast. Unpublished paper presented as part of the National Research Council workshop on "Unifying Social Frameworks: Sociocultural Data to Accomplish Department of Defense Missions," Washington, D.C., August 17, 2010.

Mirzoeff, Nicholas
2009 War is Culture: Global Counterinsurgency, Visuality, and the Petraeus Doctrine. Proceedings of the Modern Language Association of America 124 (5): 1-18.

Office of Naval Research
2010 HSCB Briefing for AFRICOM/EUCOM S&T Conference. PowerPoint briefing. Author's archives.

Panzarasa, Pietro P. and Kathleen M. Carley
2005 An Emergence-based Perspective on Organizational Culture. Unpublished paper presented at the annual meeting of the North American Association for Computational Social and Organizational Science. Notre Dame, IN, June 26-28.

Pietrocola, David
Spotlight: Barry Silverman. HSCB Newsletter. No. 3, Fall.

Price, David
2010 The Army's Take on Culture. Anthropology Now 2 (1): 57-63.

Robin, Ron
2001 The Making of the Cold War Enemy: Culture and Politics in the Military-Industrial Complex. Princeton, NJ: Princeton University Press.

Rorty, Richard
1979 Philosophy and the Mirror of Nature. Princeton, NJ: Princeton University Press.

Scott, David
2003 Culture in Political Theory. Political Theory 31 (1): 92-115.

Selmeski, Brian
2007 Military Cross-Cultural Competence: Core Concepts and Individual Development. U.S. Maxwell, AFB: Air Force Cultural and Language Center Report 2007-01.

Silverman, Barry et al.
2010 NonKin Village: An Embeddable Training Game Generator for Learning Cultural Terrain and Sustainable Counter-Insurgent Operations. In Agents for Games and Simulations. Frank Dignum et al., eds. Pp. 135-154. Berlin: Springer-Verlag.

Stockman, Farah
2010 Knowing the Enemy, One Avatar at a Time. The Boston Globe, May 30.

Turnley, Jessica Glicken and Aaron Perls
2008 What is a Computational Social Model Anyway? A Discussion of Definitions, a Consideration of Challenges, and an Explication of Process. Report No. ASCO 2008-09 for the Defense Threat Reduction Agency Advanced Systems and Concepts Office.

U.S. Army and Marine Corps
2007 Counterinsurgency Field Manual: U.S. Army Field Manual No. 3-24 and Marine Corps Warfighting Publication 3-33.5. Chicago, IL: University of Chicago Press.

U.S. Department of Defense
2010 DoD Dictionary of Military and Associated Terms. Joint Publication 1-02. (http://www.dtic.mil/doctrine/new_pubs/jp1_02.pdf), accessed January 19, 2009.

Yi, Jeonghee and Wayne Niblack
2005 Sentiment Mining in WebFoundation. Proceedings of the 21st International Conference on Data Engineering.

Zacharias, Greg L. et al., eds.
2008 Behavioral Modeling and Simulation: From Individuals to Societies. Washington, D.C: The National Academies Press.

Mark Bevir, Making Sense of Computational Models

Presented at the DTRA/Sandia Workshop on Challenges in Computational Social Science. Santa Fe, New Mexico. October 25-28, 2010.

Abstract

This paper asks: what are the implications of anti-naturalism (that is, the interpretive nature of social science) for the ways in which we conceive and use computational models? Computational models are especially commonplace as ways of generating forecasts and predictions of natural phenomena, such as the weather and earthquakes. In addition, they are widely used to aid thinking and policy-making with respect to large-scale and aggregate social phenomena and especially the unintended consequences of a vast array of actions as exemplified by traffic flows or macro-economics. Finally, a small but growing number of scholars are suggesting that computational models might play a similar role in dealing with a wider range of social and cultural activity. This suggestion has become particularly prominent in the area of intelligence and security.

It is all too easy to imagine that the differences between these domains – the natural world, unintended consequences, and social and cultural activity – are purely technical ones. Scholars often imply, for example, that the greater difficulties of modeling social and cultural activity reflect a lack of data or a lack of agreed metrics. Likewise, they imply that the more limited predictive success of models of social and cultural activity reflects the more recent origins or the more limited funding of this type of inquiry compared to, say, forecasting the weather.

Yet, there are more philosophical reasons to wonder whether models of social and cultural activity are ever likely to prove as useful as predictive tools or guides to decision-making as do models of natural phenomena. So, this paper discusses the philosophical distinctiveness of social and cultural activity and what it means for computational models.

Social and cultural activity differs from natural phenomena in that people act for reasons; they act on their beliefs and desires, whether these are conscious, tacit, or unconscious. We might say that human actions reflect the actors' interpretations of the world. When scholars explore social and cultural activity, therefore, they offer us interpretations of other people's interpretations of the world. To model natural phenomena is to interpret something, but the object or domain of the model does not embody an understanding of the world. In contrast, to model a social and cultural domain is to model actions that instantiate the actors' understanding of the world.

The interpretive nature of the social and cultural world has several implications for how we think about and use computation models of it. First, all models inevitably reflect the prior theories (or interpretations) on which they are based. In evaluating them, we need to consider the reliability of these theories, and these theories are likely to be far more contentious and far less reliable when they are theories about social and cultural activity. Second, models generally just provide data of what will happen if people act in a certain way; they do not explain why people might act as they do. To explain social and cultural phenomena, we need narratives that locate actions and the beliefs that inform them in the appropriate contexts. Finally, the contextual nature of social explanations distinguishes them from those in the natural sciences in a way that suggests they are not capable of inspiring accurate predictions.

Making Sense of Computational Models

This paper asks: what are the implications of anti-naturalism (that is, the interpretive nature of social science) for the ways in which we conceive and use computational models of different phenomena? Computational models are especially commonplace as ways of generating forecasts and predictions of natural phenomena, such as the weather and earthquakes. In addition, they are widely used to aid thinking and policy-making with respect to large-scale and aggregate social phenomena and especially the unintended consequences of a vast array of actions as exemplified by traffic flows or macro-economics. Finally, a small but growing number of scholars are suggesting that computational models might play a similar role in dealing with a wider range of social and cultural activity. This suggestion has become particularly prominent in the area of intelligence and security.

It is all too easy to imagine that the differences between these domains – the natural world, unintended consequences, and social and cultural activity – are purely technical ones of the sort David Sallach discusses in his paper. Scholars often imply, for example, that the greater difficulties of modeling social and cultural activity reflect a lack of data or a lack of agreed metrics. Likewise, they imply that the more limited predictive success of models of social and cultural activity reflects the more recent origins or the more limited funding of this type of inquiry compared to, say, forecasting the weather. While such technical concerns may be relevant, there are also broader philosophical debates of direct relevance to the way in think of computational models of different types of phenomena. It is these philosophical debates that I will discuss in this paper.

The paper begins, in its first section, with a discussion of one prominent philosophical view of social science – anti-naturalism – and the challenges it poses to social science. Anti-naturalists emphasize the philosophical differences between the natural and social sciences that follow from the intentional nature of human action. Next, in the second section of the paper, I explore the relevance of philosophical views such as anti-naturalism to methodological debates in social science. While the

discussion makes the case for bringing philosophical debates to bear on questions about techniques such as computational modeling, but it also tries to specify the role that philosophy might play in evaluating techniques such as computation modeling. Finally, in the third section of the paper, I consider the implications of anti-naturalism for the way we think about computational models of human cultural and social behavior.

Naturalism and Anti-Naturalism

A prominent faultline that continuously surfaces in debates over computational and other methods in social science is that between positivist or scientific approaches on the one hand, and postpositivist or critical approaches on the other (1998; Dahl et al. 2004; Tickner 1998; Oren 2006; Johnson 2006; Marsh and Savigny 2004). Alas, these debates are often conducted with little reflection on the philosophical underpinnings of the relevant approaches. In these debates, concepts like "positivist" and "postpositivist" are often associated with methodological choices – quantitative or qualitative – at least as much as philosophical commitments – naturalism or anti-naturalism. The resulting lack of philosophical reflection can result in a skewed understanding of the issues at stake in the debates over an adequate social science. One plausible example is Robert Albro's suggestion, in his paper, that the military treats culture as a natural system because of their prior frames of reference and without due reflection.

One problem with the usual faultline between positivism and postpositivism is the unhelpful way in which it muddles methodological and philosophical concerns. The term "positivism" is often used to fuse foundationalism and naturalism with quantitative methods. And the opposite term, "postpositivism", is made to fuse postfoundationalism and anti-naturalism with qualitative and interpretive methods. Yet, things are not that simple. Some proponents of quantitative methods are naturalists who have doubts about positivist epistemologies. And some proponents of qualitative

methods appear to embrace a rather naïve empiricism in justifying their methods as ways of getting at facts that elude quantitative scholars. Perhaps we might break with the usual faultline between positivism and postpositivism in order to focus initially on philosophical issues, specifically the key distinction between naturalism and anti-naturalism.

The dazzling achievements of the natural sciences have exerted an enormous pressure on the social sciences, including a powerful drive to model the latter on ontological and epistemological foundations associated with the former. Naturalism arises from the belief that similarities between the natural and social worlds are such that they should be studied in the same ways. Although naturalists might suggest, as Sallach does in his paper, that social systems are dramatically different from natural systems, they still believe that the same logic of inquiry applies to both. Initially, naturalists wanted mainly to preclude appeals to supernatural explanations: they argued that humans were part of nature and so amenable to empirical study, and they promoted a scientific method based on the rigorous collection and sifting of facts. Before long, however, naturalism became ensnared with the positivist conviction that the same logic of inquiry applies to both the natural and social sciences. Hence, today we can define naturalism as the idea that the social sciences should strive to develop predictive and causal explanations akin to those that are found in the natural sciences (Ayer 1967; Hempel 1942). In the classic statements of this view, the social sciences study fixed objects of inquiry that possess observable and, at least to some extent, measurable properties, such that they are amenable to explanations in terms of general laws, even if these general laws sometimes involve assigning probabilities to various outcomes.

In the past few decades, however, philosophers of social science have typically come to favor anti-naturalism. The critique of naturalism has developed over the past half-century within a variety of philosophical traditions. Anti-naturalism has been most clearly and consistently articulated in the hermeneutic tradition, starting with the work of Wilhelm Dilthey (1976) at the turn of the

twentieth century and developed more recently by Hans-Georg Gadamer (2002) and Paul Ricoeur (1976). In the social sciences, Max Weber (1978) was one scholar who incorporated some hermeneutic themes. He insisted that causal explanation in social science relied in large part on verstehen (interpretive understanding). And he insisted on the singularity of social explanation; they seek contextually specific causes of historical particulars. Nowadays anti-naturalism is also dominant in analytic philosophy. Its dominance therein began in the latter half of the twentieth century following the leads provided by Ludwig Wittgenstein (2001),[43] Alasdair MacIntyre (1969), and Charles Taylor (1971). Yet more contributions to the rise of anti-naturalism have come from phenomenology (Husserl 1970), pragmatism (Dewey 1960; Rorty 1980), ethnomethodology (Garfinkel 1967), and cultural anthropology (Geertz 1973).

Anti-naturalists argue that constitutive features of human life set it apart from the rest of nature to such an extent that the social sciences cannot operate as do the natural sciences. Whereas naturalists such as Sallach suggest these features make social science more complex than natural science, anti-naturalists insist that they make the model of the natural sciences inappropriate for the study of human social and cultural behavior. The relevant features of human action are that it is meaningful and historically contingent. Let us explore them in turn before then emphasizing that they apply as much to social scientists as to those who they study.

I will begin with the meaningfulness of social life. Some naturalists hold a positivist epistemology according to which causal explanations are validated by their fit with observations, and meanings are irrelevant because they are not observable. These positions informed, for instance, classical behaviorism as propounded by John B. Watson (1924) and B. F. Skinner (1938). However, because this positivist epistemology is rarely espoused nowadays, I will concentrate on naturalists who would agree that human actions have meanings for those who perform them. It is widely

[43] Wittgenstein has inspired numerous studies of the interpretive nature of social science. Examples include: Bevir 1999; Pitkin 1972; Winch 1958.

accepted today that agents act for reasons of their own, albeit that we sometimes take the reasons to be tacit, subconscious, or even unconscious, as opposed to explicit and conscious. What divides naturalists and anti-naturalists is the role they give to meanings in the explanation of actions and so of the explanation of social and cultural patterns arising from actions.[44] Naturalists typically want meanings to drop out of these explanations. Philosophical exponents of naturalism argue, for instance, that to give the reasons for an action is merely to re-describe that action. If we want to explain an action, they add, we have to show how it – and so no doubt the reason for which the agent performed it – conforms to a general law couched in terms of social facts (Ayer 1967).

Anti-naturalists refuse to let meanings and beliefs drop out of explanations in the human sciences. They argue that meanings are constitutive of human action. Hence, as Clifford Geertz (1973: 5) claimed, social science needs to be "not an experimental science in search of law but an interpretive one in search of meaning." Anti-naturalists here uphold the centrality of meanings for social science on the grounds not only that actions are meaningful but also that these meanings are holistic. In this view, we can properly understand and explain people's beliefs only by locating them in a wider context of meanings. Meanings cannot be reduced to allegedly objective facts since their content depends on their relationship to other meanings. The social sciences require a contextual form of explanation that distinguishes them from the natural sciences. So, anti-naturalism points to the importance of elucidating and explaining meanings by reference to wider systems of meanings, rather than by reference to categories such as social class or institutional position, and rather than by construing ideas or meanings as "independent variables" within the framework of naturalist forms

[44] For our current purposes, we might put to one side the question of whether understanding is or is not a species of explanation. The dubious relevance of this question appears in the fact that naturalists and anti-naturalists alike are divided upon it. Indeed, I suspect the question is just a terminological one. When naturalists or anti-naturalists deny that understanding is a species of explanation, they are identifying explanation with the causal explanations found in the natural sciences. When they allow that understanding can be a type of explanation, they are adopting a broader concept of explanation (and perhaps also cause) such that to explain something is just to say why it is as it is. Perhaps the most insightful discussion of this issue is that by Donald Davidson. In a series of essays, Davidson (1980) argued that reasons were the causes of actions, that the relevant concept of cause was that found in our folk psychology, and that these causes might map onto physical causes of which we as yet did not have secure knowledge.

of explanation. It is worth adding that the dominance of meaning holism in contemporary philosophy – as observed even by skeptics (Fodor and LePore 1992) – suggests that naturalism might prove a difficult doctrine for political scientists to defend.

Let me turn now to the historically contingent nature of human action. When naturalists try to let meanings drop out of their explanations, they are usually hoping at least to point toward classifications, correlations, or other regularities that hold across various cases. Even when they renounce the ideal of a universal theory or law, they still regard historical contingency and contextual specificity as obstacles that need to be overcome in the search for cross-temporal and cross-cultural regularities. Greg Luebbert, for example, discusses a number of discrete national case studies but his ultimate aim is to find "a single set of variables and logically consistent causal connections that make sense of a broad range of national experiences" (Luebbert 1991: 5). Naturalists typically search for causal connections that bestride time and space like colossi. They attempt to control for all kinds of variables and thereby arrive at parsimonious explanations. But they can do so only by freezing history.

In stark contrast, anti-naturalists argue that the role of meanings within social life precludes regularities standing as explanations. That said, we need to be careful how we phrase what is at issue here. Anti-naturalists have no reason to deny that social scientists might offer general statements that cover diverse cases. Rather, they typically object to two specific features of the naturalist view of generalizations. First, anti-naturalists deny that general statements constitute a uniquely appropriate or powerful form of social knowledge. They consider statements about unique and contingent aspects of particular social phenomena to be at least as apposite and valuable as general statements. In their view, generalizations can deprive our understanding of social phenomena of what is most distinctly and significantly human about them. Second, anti-naturalists reject the claim that general statements can provide explanations of features of particular cases: just as we can say that X, Y, and

Z are all red without explaining anything else about them, so we can say that X, Y, and Z are all democracies but that does not explain any other feature they might have in common. Anti-naturalists oppose explanations of human actions in terms of trans-historical generalities because they conceive of human action as inherently contingent and particular. Human life is characterized by contingency, temporal fluidity, and contextual specificity. Hence we cannot explain social phenomena adequately if we fail fully to take into account both their inherent flux and their concrete links to specific contexts. The social sciences require a historical and contingent form of explanation that distinguishes it from the natural sciences.

So, anti-naturalists emphasize the meaningful and contingent nature of human life. These emphases apply to social scientists as much as to the people they study. Social scientists come to hold particular beliefs against the background of contingent traditions. Naturalists may treat the situatedness of the social scientist as an obstacle to be overcome in the pursuit of proper knowledge. They ask social scientists to try to abstract themselves from their historical perspectives. They want social science knowledge to present itself as divested of particular theories and prejudices. In contrast, anti-naturalists usually deny the very possibility of abstracting ourselves from our prior webs of belief. They suggest that social science always takes place from within particular linguistic, historical, and ethical standpoints. They questions asked and the concepts formed by social scientists are always informed by their existing webs of belief and their assumptions.

The combined recognition of, on the one hand, the situatedness of the social scientist and, on the other hand, the meaningfulness of social life introduces a dialogical dimension to social science. Naturalists typically construe explanation as the product of a unidirectional subject-object relationship. Their neglect of the constitutive role of meanings leads them to see the social scientist as the only agent involved in crafting explanations: the objects of social science are just that – passive objects to be studied. In contrast, anti-naturalists often conceive of explanation as the

product of a kind of dialogue between social scientists and those they study. Social science generally involves a subject-subject interaction in which the scholar responds to the interpretations or meanings of the relevant social actors. An encounter with the beliefs or meanings of social actors always has the potential to send out ripples through a scholar's own beliefs, altering their understanding of, say, their research agendas, the traditions in which they work, or their normative commitments.

Philosophy and Method

Anti-naturalists argue that: actions are meaningful, meanings are contingent and liable to change over time, and these facts apply to the actions and beliefs of social scientists as well as those whom they study in a way that points to dialogical modes of inquiry. What are the implications of this anti-naturalism for the way we think about computational models? As I mentioned earlier, anti-naturalism is often confusingly fused with a commitment to qualitative methods. It should not be. There is no logical reason to fuse positivism with quantitative methods and anti-naturalism with qualitative methods. Indeed, more generally still, no philosophical position legislates strictly for or against the use of any methodological technique.

To understand the implications of anti-naturalism for computational models, we must first grasp the relationship of philosophy to methodology. The problem here is that approaches to social science are strange beasts. Most contain a jumble of philosophical theories, methodological techniques, and empirical topics. While the relevant theory, techniques, and topics may have some ties to one another, they definitely do not logically entail each other. For example, social scientists often talk about behavioralism as if it were a coherent whole, but really there are no necessary ties between positivist theory, large-N statistical techniques, and behavioral topics.

While philosophical theories do not legislate for or against particular techniques, it seems far more reasonable to expect them to help us to understand the nature of the data that techniques generate and so to help us judge the appropriateness of any given technique for the problems social scientists address. Whether we call it "philosophy" or something else such as "meta-methodology", we should promote steady and deliberate theoretically reflection on what methods are appropriate to the study of what aspects of social and cultural behavior. Philosophy helps to clarify what kind of knowledge and what kind of explanations fit the kinds of objects that are the concern of social science.

Anti-naturalism does not limit social scientists to textual readings and small-scale observations; it does not exclude survey research, quantitative studies, and computational models. To the contrary, anti-naturalists can construct their interpretations using data generated by various techniques. They can draw on participant observation, interviews, questionnaires, mass surveys, statistical analysis, and computational models as well as reading memoirs, newspapers, and official and unofficial documents. Anti-naturalism does not prescribe a particular methodological toolkit for producing data. Instead, it prescribes a particular way of treating data of any type. It implies that social scientists should treat data in ways consistent with the meaningful and contingent nature of human action. They should treat data as evidence of the historically-situated beliefs embedded in actions and practices.

The importance of philosophy to methodology should now be clear. Only when we know what kinds of knowledge and explanation are apt for social science can we intelligently decide what methods are best suited to producing them and what any given method has to contribute. Whether we believe any method to be apt in any given instance necessarily depends on our underlying philosophical views. The only question is whether we consciously aware of these views and so open to trying to improve them, or whether we risk being confused. Again, social scientists should not let

the undoubted importance of methodological rigor obscure what are prior philosophical issues about the adequacy of the commitments entailed by any claim that any method is an appropriate means of generating knowledge about any given type of object. Discussion of methods and their utility are profoundly impoverished by a lack of reflection on the relevant philosophical assumptions. Many social scientists have worried about hyperfactualism – the collection of data without proper theoretical reflection. Today, they might worry at least as much about hypermethodologism – the application of methodological techniques without proper philosophical reflection.

What, then, are the philosophical implications of anti-naturalism for the way in which social scientists should think about data and models? I will highlight three main implications, loosely corresponding respectively to the anti-naturalist emphases on the situatedness of the scholar, the meaningfulness of action, and historical contingency.

Consider, first, the situatedness of the scholar. Anti-naturalism suggests that claims to knowledge are inherently theory-laden. People's accounts of the world are never straightforwardly verifiable or falsifiable by reference to allegedly given facts. What would have to be the case for a proposition to be true (or false) depends on the other propositions we hold true. People can logically reject or retain any proposition in the face of any evidence provided they make appropriate changes to other propositions they hold true. No proposition ever confronts the world in splendid isolation. Evidence only ever confronts overarching webs of belief, and even then the evidence is saturated by theories that are part of the relevant webs of belief. Fortunately to insist on the theory-laden nature of knowledge is not to say it is groundless. Many philosophers offer alternative accounts of justified knowledge based on comparative approaches to theory choice. They argue that a theory or other belief is justified because it is better than the others available to us. For my purposes now, however, the important point is that no method – computational models, regression

analyses, etc. – can conclusively justify the explanatory claims, predictions, or data they generate. Methods just create data, the validity of which is still open to debate. The validity of both data and causal claims depends on comparisons between rival bundles of facts, theories, and assumptions.

A second implication of anti-naturalism derives from its emphasis on the meaningful nature of action. Anti-naturalism leads here to a linguistic constructivism according to which people not only make the social world by their actions; they also make the meanings and beliefs on which we act. People's beliefs, concepts, actions, and so practices are products of particular traditions or discourses. Social concepts (and social objects), such as "bureaucracy" or "democracy", do not have intrinsic properties and objective boundaries. They are artificial inventions of particular languages and societies. Their content varies with the wider webs of belief in which they are situated. Crucially, this linguistic constructivism implies that social concepts rarely (if ever) refer to natural kinds. It undermines attempts to ascribe to social objects an essence that determines their other properties and the effects they have. Linguistic constructivism implies that social concepts are pragmatic. Social life consists of meaningful activity. When social scientists use aggregate concepts to refer to a set of actions, the decision about which actions to include under the concept is a pragmatic one made in accord with their purposes.

The third implication of anti-naturalism stems from recognition of the historical contingency of beliefs and actions. For a start, because people act on webs of belief, social scientists can properly explain people's beliefs (and so actions) only by locating them in the context of the relevant web. Social explanations should elucidate beliefs by showing how they relate to one another, not by trying to reduce them to categories such as social class or institutional position. In addition, the historical contingency of these webs of belief implies that social scientists cannot explain why people hold the webs of belief they do solely by reference to people's experiences, interests, or social location. To the contrary, even people's beliefs about their experiences, interests, and social location will depend

on their prior theories. Thus, a social scientist can explain why people hold the webs of belief they do only by reference to the intellectual traditions that these people inherit. Even the concepts, actions, and practices that seem most natural to us need to be explained as products of a contingent history. Social explanation contains an inherently historicist moment. This historicism means, finally, that correlations, classifications, and models are not properly speaking explanations at all. They are just further types of data that we will accept in so far as we trust the methods by which they are produced. Social scientists can explain such data only by appealing to contextual and historical narratives.

Interpreting Computational Models

Anti-naturalism has several implications for the ways in which we should think about data and models. In the rest of this paper, I want to point to some of more specific implications for the roles that computational models might play in policy-making. Here too I highlight three main implications, loosely corresponding respectively to the anti-naturalist emphases on the situatedness of the scholar, the meaningfulness of action, and historical contingency.

First, all models inevitably reflect the prior theories (or interpretations) on which they are based. When policy makers use or evaluate computational models, they should consider the reliability of these theories. Of course, all models, whether of natural or social phenomena rest on theories that are open to question. However, the theories embedded in models of social and cultural behavior are likely to be far more contentious and far less reliable than those embedded in models of natural phenomena. The reasons for being more suspicious of social theories are, moreover, as much philosophical as they are practical.

In particular, models of human activity almost inevitably depend on theories that reify human activity in a way that we know is false. These models almost inevitably lead to simplifications,

albeit simplifications that might sometimes aid decision-making. Here, models of cultural and social behavior generally require theoretical assumptions that reify activity by implying that people of a certain type or people in certain situations will act in a given way. That is to say, these models embody theories that suggest certain objective social facts will reliably lead to the persistence of patterns of action. The theories behind these models highlight or postulate patterns among actions and beliefs only by obscuring differences that may be or may become extremely important in other contexts or in later settings. The worry is that if policy makers overly privilege the models, they may think – as Jessica Turnley in her paper suggests they do – that the models are uniquely correct descriptions of the world; they may ignore the complexities and diversities that the models obscure, or the possibility that such complexities and diversities may arise in the future.

In particular, when a model projects a pattern of human activity, policy makers should generally consider whether the model hides differences in the nature of the action and the beliefs inspiring the action. They should consider the possibility that different beliefs just happen to have produced similar actions, or that different webs of belief just happen to have shared some similar features. Some patterns arise when people act in similar ways for very different reasons. For example, there is a well-established (if rapidly declining) correlation in Britain between being working class and voting for the Labor Party. The worry is that this correlation may lead modelers and policy makers to think in terms of a monolithic pattern of activity. Yet, different working-class people vote Labor for different reasons. Some may vote Labor because they believe that they are working class and Labor will promote the interests of the working class. Others may vote Labor because they think they are working class, do not think Labor will promote the workers' interests, but nonetheless identify emotionally with the symbolism of the Labor Party. Others may believe (perhaps mistakenly) that they are middle class and yet vote Labor because they see themselves as committed to values such as social justice. Patterns also arise – especially in speech, beliefs, and

attitudes – because people have webs of belief that have some abstract features in common but are very different in their specifics. For example, suppose that many people from a particular nationality or religion say they support a strong state. Modelers and policy makers might think that they have found a monolithic pattern. They should be aware, however, that different members of the group may mean different things when they use the word "state" or may have very different reasons for advocating a strong state. Some may think their state is unable to defend the rule of law and just want it to do so. Others may want the state to impose stronger moral norms on society. The general point is, of course, that whenever a model hides differences that would appear if we asked whether different beliefs happen to have produced similar actions, or whether different webs of belief happen to include some similar features, then the model is likely to be particularly misleading as a guide to policy.

Like the situatedness of the scholar, the meaningfulness of action has implications for the roles that computational models might play in policy-making. Policy makers should remember that correlations and the like are not explanations. Explanations in the social sciences require narratives that reveal the meaningful nature of action and that relate meanings and beliefs to one another in webs. While computational models can provide data or insights that contribute to such narratives, policy makers should not become too preoccupied with them or too reliant upon them. Instead, policy makers should recognize that computational models are just further data to be included in a narrative, or at most themselves narratives about how people have acted or will react given their beliefs and desires. No matter what rigor or expertise modelers bring to bear, all models we can do is tell a story and judge what the future might bring.

To explain social and cultural phenomena, policy makers need narratives that locate actions and the beliefs that inform them in the appropriate contexts. Historical and fictional narratives characteristically relate actions to the beliefs and desires that produce them. Narratives depend here

on conditional connections that are not necessary or arbitrary: it is because they are not necessary that social science differs from the natural sciences, and yet it is because they are not arbitrary that social scientists can use them to explain actions and practices. These conditional connections exist when the nature of one object draws on the nature of another. The relevant objects condition each other, so they do not have an arbitrary relationship. But neither object follows inexorably from the other, so they do not have a necessary relationship. Social knowledge depends on telling stories that postulate just such conditional connections between beliefs, actions, practices, and their contexts. Practitioners should experiment with multiple stories that reveal new aspects of situations; they should hear different voices, talk to one another, and develop tentative and evolving narratives. Even when computational models provide data on possible outcomes, policy-makers should experiment with different ways of narrating these outcomes, and they should explore alternative stories – based on alternative theories – that suggest different outcomes. One valuable role of models might be precisely to increase the range of stories that analysts and decision-makers consider.

Indeed, I would go a step further. The meaningfulness of action suggests that policy makers would be well advised at least to supplement computational models with data that provides a richer picture of actors' beliefs and the ways in which these beliefs coalesce in webs. Here, anti-naturalism might have some heuristic implications for data collection, prompting a greater emphasis on qualitative methods. For example, suppose that the data provided by comparing and codifying the formal constitutional documents of democracies leads social scientists to attribute certain beliefs to their political leaders and then to build models based on the prior theory that democratic leaders hold these beliefs.[45] As we have seen, because the data encourages social scientists to reify common

[45] I am thinking of the literature on the idea of a democratic peace. For a rather partisan bibliography see http://www.hawaii.edu/powerkills/BIBLIO.HTML. For an example of the way anti-naturalism might prompt skepticism, see Oren (2003).

patterns, it may elide differences between people. Thus, social scientists and policy makers might consider supplementing such data and models with detailed studies of the beliefs of the relevant people using methods such as textual analysis, participant observation, and in-depth interviews. Much contemporary social science ignores, or even denigrates such methods, preferring abstract models. While anti-naturalism does not require the use of any type of data or method, it does redress the balance, reminding social scientists and policy makers of the value of the qualitative analyses at least as a supplement to other types of data and methods.

Let us turn, finally, to the implications of contingency for the roles that computational models might play in policy-making. Anti-naturalism suggests that human activity is a series of contingent (perhaps even accidental) appropriations, modifications, and transformations from the old to the new. Again, change occurs contingently as people reinterpret, modify, or transform an inherited tradition in response to novel circumstances or other dilemmas. Moreover, this contingency implies that human life is radically open in that what happens is always contestable. There are always innumerable ways in which an action, practice, or traditions may be reinterpreted, transformed, or overpowered. We should be suspicious of attempts to portray any social practice as unified or change as based on a uniform consensus.

Crucially, the contingent nature of social life undermines the very possibility of prediction – defined in contrast to the looser idea of informed conjecture. The contextual and historical nature of social explanations precludes there giving rise to the kinds of accurate predictions associated with the natural sciences. Because social phenomena are not natural kinds, social scientists can not explain or predict social life by appealing to uniform laws. Social life is a product of the ways in which people act so as to modify inherited traditions and practices, and the ways in which they do so are open-ended and so not amenable to prediction. Because traditions and practices are not fixed, we cannot know in advance how people will develop their beliefs and actions in response to any new

situation or dilemma. Therefore, a social scientist cannot strictly predict how behave, no matter what they know about their past behavior or what they know about how similar people have behaved. Whatever limits social scientists build into their predictions, people could arrive at new beliefs and actions outside those limits. Social scientists cannot make predictions. All they can offer are informed conjectures that seek to explain practices and actions by pointing to the conditional connections between actions, beliefs, traditions, and dilemmas. Their conjectures are stories, understood as provisional narratives about possible futures.

It is true that some naturalists have attempted to rebuff anti-naturalism here by equating social science with the study of systems or structures that cannot be understood as the intended consequence of a single action. Traffic jams are often evoked as examples of such structures. But traffic jams and other such structures scarcely undermine anti-naturalism. Most of what we want to know about traffic jams comes down to intentional action. To explain why people are driving when and where they are, we want to know whether they intend (consciously or not) to go to work, to a sports game, shopping, visiting relatives, and so on. Even more generally, we might explore the wider webs of belief that constitute the social practices within which these intentions are embedded. Why do people believe that driving to work is better than using public transportation? Why don't they take political action to increase public investment in transportation infrastructure? All such questions are questions about meaningful intentionality. If an account of traffic jams or other such structures really did ignore intentionality, it would be a very thin and inadequate account. It could tell us only in purely physical terms that the traffic jam arose because a given number of people tried to drive cars along a stretch of road of given dimensions. It could tell us nothing about the actions that led to these physical consequences; it could not tell us why these people were driving their cars or why the road system is as it is.

Conclusion

My main aim in this paper has been to raise awareness of the importance of philosophical issues for our understanding of the nature, use, and limits of computational models. The references include a number of the most important philosophical so that the reader can further explore and reflect on these issues. It seems to me that the majority of philosophers today hold to some form of the anti-naturalist view that human life differs from the natural world in that humans have a capacity for acting for reasons of their own. It also seems to me that this anti-naturalism reflects the common-sense of our age. After all, I expect that if I asked you why you are attending this workshop, you would answer me by giving your reasons – you would refer to your beliefs and desires and their place in a historical narrative about your life and its context. My other main aim has thus been to explore the implications of anti-naturalism for our understanding of the nature, use, and limits of computational models. I have suggested here that there are philosophical reasons to wonder whether models of social and cultural activity are ever likely to prove as useful as predictive tools or guides to decision-making as do models of natural phenomena. The point here is not that anti-naturalism repudiates computational models (or, for that matter, any other methodological technique); it does not. The point is, rather, that anti-naturalism undermines some of the assumptions that are widespread among scholars attempting to develop computational models of human social and cultural behavior. Anti-naturalism suggests, more specifically, that: i) models rely on theories that are almost always overly-simplified reifications; ii) the reified patterns embedded in and offered by models can be properly explained only by narratives; and iii) models can never guide decisions by offering predictions; they are at most conjectures about possibilities that policy makers might take into account in making a decision.

Given that much of my argument has been general, it is important to bring it down to earth with a thud. Most policy advisers will accept that the arts of storytelling and decision-making are

integral parts of their work. Practitioners often use phrases such as: "Have we got our story straight?", "Are we telling a consistent story?", and "What is our story?". Advisors often explain past events to justify recommendations for the future. Anti-naturalism makes sense of the kind of knowledge they are seeking and using to make decisions. In short, a stress on interpretation and storytelling is not an example of academic whimsy. It reminds policy-makers of what they do, and explains why doing that remains a valuable corrective to an over-reliance on formal methods and techniques such as computational models.

Bibliography

Ayer, A.J. 1967. Man as a Subject for Science. In Philosophy, Politics and Society, 3rd Series, ed. P. Laslett and W. Runciman. Oxford: Basil Blackwell.

Bevir, Mark. 1999. The Logic of the History of Ideas. Cambridge: Cambridge University Press.

Davidson, Donald. 1980. Essays on Actions and Events. Oxford: Clarendon Press.

Dewey, John. 1960. The Quest for Certainty: A Study of the Relation of Knowledge and Action. New York: Putnam.

Dahl, Robert A., et al. 2004. What Have We Learned? In Problems and Methods in the Study of Politics, ed. I. Shapiro et al. Cambridge: Cambridge University Press, 377-394.

Dilthey, Wilhelm. 1976. Selected Writings, ed. and trans. H. Rickman. Cambridge: Cambridge University Press.

Fodor, Jerry, and Ernest LePore. 1992. Holism: A Shopper's Guide. Oxford: Blackwell.

Gadamer, Hans-Georg. 2002. Truth and Method. New York: Continuum.

Garfinkel, Harold. 1967. Studies in Ethnomethodology. Englewood Cliffs, NJ: Prentice-Hall.

Geertz, Clifford. 1973. The Interpretation of Cultures: Selected Essays. New York: Basic Books.

Hempel, Carl. 1942. The Function of General Laws in History. Journal of Philosophy 39: 35-48.

Husserl, Edmund. 1970. The Crisis of European Sciences and Transcendental Phenomenology: An Introduction to Phenomenological Philosophy, tr. D. Carr. Evanston, IL: Northwestern University Press.

Johnson, James. 2006. Consequences of Positivism: A Pragmatist Assessment. Comparative Political Studies 39(2): 224-52.

Lin, Ann Chih. 1998. Bridging Positivist and Interpretivist Approaches to Qualitative Methods. Policy Studies Journal 26(1): 162-180.

Luebbert, Gregory M. 1991. Liberalism, Fascism, or Social Democracy: Social Classes and the Political Origins of Regimes in Interwar Europe. New York: Oxford University Press.

MacIntyre, Alisdair. 1969. A Mistake about Causality in Social Science. In Philosophy, Politics and Society, 2nd Series, ed. P. Laslett and W. Runciman. Oxford: Basil Blackwell.

Marsh, David, and Heather Savigny. 2004. Political Science as a Broad Church: The Search for a Pluralist Discipline. Politics 24(3): 155-168.

Oren, Ido. 2003. Our Enemies and US: America's Rivalries and the Making of Political Science. Ithaca: Cornell University Press.

Oren, Ido. 2006. Can Political Science Emulate the Natural Sciences? The Problem of Self-Disconfirming Analysis. Polity 38(1): 72-100.

Pitkin, Hannah F. 1972. Wittgenstein and Justice: On the Significance of Ludwig Wittgenstein for Social and Political Thought. Berkeley: University of California Press.

Ricoeur, Paul. 1976. Interpretation Theory: Discourse and the Surplus of Meaning. Fort Worth: Texas Christian University Press.

Rorty, Richard. 1980. Philosophy and the Mirror of Nature. Princeton: Princeton University Press.

Skinner, Burrhus F. 1938. The Behavior of Organisms: An Experimental Analysis. New York: Appleton-Century.

Taylor, Charles. 1971. Interpretation and the Sciences of Man. Review of Metaphysics 25(1): 3-51.

Tickner, J. Ann. 1998. International Relations: Post-Positivist and Feminist Perspectives. In New Handbook of Political Science, ed. R. E. Goodin and H.-D. Klingemann. Oxford: Oxford University Press, 446-462.

Watson, John B. 1924. Behaviorism. New York: Norton.

Weber, Max. 1978. Economy and Society, 2 vols., ed. G. Roth and C. Wittich. Berkeley: University of California Press.

Winch, Peter. 1958. The Idea of a Social Science. London: Routledge & Kegan Paul.

Wittgenstein, Ludwig. 2001. Philosophical Investigations. Malden, Mass.: Blackwell.

Francois M. Hemez, Is "Predictability" in Computational Sciences a Myth?

Presented at the DTRA/Sandia Workshop on Challenges in Computational Social Science. Santa Fe, New Mexico. October 25-28, 2010.

Abstract:

Within the last two decades, Modeling and Simulation (M&S) has become the tool of choice to investigate the behavior of complex phenomena. Successes encountered in "hard" sciences are prompting interest to apply a similar approach to Computational Social Sciences in support, for example, of national security applications faced by the Intelligence Community (IC). This manuscript attempts to contribute to the debate on the relevance of M&S to IC problems by offering an overview of what it takes to reach "predictability" in computational sciences.

Even though models developed in "soft" and "hard" sciences are different, useful analogies can be drawn. The starting point is to view numerical simulations as "filters" capable to represent information only within specific length, time or energy bandwidths. This simplified view leads to the discussion of resolving versus modeling which motivates the need for sub-scale modeling. The role that modeling assumptions play in "hiding" our lack-of-knowledge about sub-scale phenomena is explained which leads to discussing uncertainty in simulations. It is argued that the uncertainty caused by resolution and modeling assumptions should be dealt with differently than uncertainty due to randomness or variability. The corollary is that a predictive capability cannot be defined solely as accuracy, or ability of predictions to match the available physical observations. We propose that "predictability" is the demonstration that predictions from a class of "equivalent" models are as consistent as possible. Equivalency stems from defining models that share a minimum requirement of accuracy, while being equally robust to the sources of lack-of-knowledge in the problem. Examples in computational physics and engineering are given to illustrate the discussion.

1. Introduction

With the convergence, in the past two decades, of TeraOPS46 computing, cheap memory, mature algorithms, fast-access databases and seamless graphics, "hard" sciences have turned to Modeling and Simulation (M&S) to investigate the behavior of complex systems. At the U.S. Department of Energy and since 1995, the Advanced Scientific Computing (ASC) Program has been tasked with the development and implementation of the predictive capability required to assess the performance, safety and reliability of our nuclear deterrent without resorting to full-scale testing [1]. Other examples of M&S programs that support high-consequence decisions include modeling the evolution of global climate; understanding the effects of a terrorist nuclear explosion in urban environment; and proposing efficient response scenarios to face epidemic outbreaks. In industry, manufacturers routinely resort to numerical models to design engineered systems, and perform performance certifications or safety assessments such as flutter stability in the aerospace industry or crash worthiness in the automotive industry [2]. These successes have, more recently, prompted interest within the Intelligence Community (IC) that is expressing interest in the capability to

[2] One TeraOPS is equal to 10^{+12} floating-point operations (such as a multiplication) per second.

understand social trends and predict behaviors through M&S with the ultimate goal of better anticipating threats to our national interests.

This manuscript attempts to contribute to the debate on the relevance of M&S to IC applications and, more broadly, the emergence of Computational Social Science (CSS), by offering a brief overview of what it takes to reach "predictability" in computational sciences. Even though "soft" sciences study phenomena and develop models that are different from those of disciplines such as, for example, computational physics, useful analogies can be drawn. Our goal is to propose a practical definition of "predictability" so that it does not turn into an unattainable myth.

In this work, the numerical simulations are viewed at a conceptual level as "filters" that represent information only within specific length scales, time scales or energy bandwidths. This simplified view leads to the discussion of resolving versus modeling. Performing numerical simulations capable of resolving all phenomena of interest is clearly not possible because it would ultimately face fundamental barriers and uncertainty principles such as those put forth by Heisenberg in 1927. To alleviate the lack-of-resolution in numerical simulations, sub-scale modeling becomes a necessary "evil." While bringing closure to the modeling effort, the introduction of a feedback mechanism to account for sub-scale information introduces arbitrary choices and assumptions. Two important questions of "predictability" become, first, assessing the extent to which these assumptions are warranted; and, second, understanding their effects on predictions.

These questions lead to the need to generalize the concept of uncertainty to numerical models. In the context of M&S, uncertainty can be described using metrics of entropy, a concept that was first introduced by Clausius in 1865. Because the origin and nature of this uncertainty may be different from conventional variability and randomness, it is argued that assessing predictive capability solely in terms of prediction accuracy makes little-to-no sense. How valuable is it, for example, to demonstrate that predictions are 3% accurate when some of the models used in the numerical simulation are based on unwarranted or, worse, incorrect assumptions?

We propose, instead, to define "predictability" based on understanding the trade-offs between fidelity-to-data, consistency of numerical predictions for classes of "equivalent" models, and robustness to lack-of-knowledge. Based on a broad definition of uncertainty that introduces little-to-no practical limitation, we demonstrate that the fidelity-to-data and robustness of a model cannot be simultaneously improved. Likewise, it is shown that increasing robustness comes at the expense of making less consistent predictions with a class of equivalent models. It is argued that assessing these trade-offs, and communicating them efficiently to stakeholders, is precisely the mechanism by which numerical simulations can support decision-making. Believing, on the other hand, that a predictive capability can be achieved by relying on the calibration of models to the available physical observations is nothing but a myth. The definition of "predictability" that we arrive at is not based solely on accuracy, even though it remains an essential ingredient. We propose that "predictability" is the ability to make as-consistent-as-possible predictions from a class of equivalent models. Equivalency stems from the fact that all models included in the class share a minimum requirement of accuracy while being equally robust to the sources of lack-of-knowledge in the problem.

The significance of the above definition is that it shifts the problematic away from the sterile, yet, often encountered, discussion of which model is "correct." We argue that it does not matter to know which model may be more appropriate. What is important is to show that, no matter which option is considered, the predictions made with equivalent models are consistent. Confidence

ultimately comes from showing that a class of equally accurate and equally robust models, that may include very different models and numerical options, produces consistent predictions.

The manuscript is organized as follows. Section 2 addresses what it may take to be "predictive." The trade-offs between resolving the information content of a given problem and modeling it are discussed in section 3. The third section also emphasizes that the need to model the unresolved information unavoidably introduces modeling uncertainty, which leads to assumption-making. In section 4, uncertainty in computational sciences is discussed with focuses on, first, exploring the diversity of uncertainty encountered and, second, briefly explaining the implication that lack-of-resolution has on numerical predictions. Finally, section 5 illustrates the concept of trading-off accuracy, robustness to lack-of-knowledge, and consistency of predictions using an application to the verification of a computer code.

2. A Brief Discussion of What Makes a "Predictive" Simulation

There are two basic drivers for the development of models in computational sciences and engineering. One objective can be to describe a phenomenon without necessarily attempting, or needing, to understand it with precision. This would be the case when one does not need to understand where the phenomenology originates from, how it "works," how repeatable it may be, and what are the actions that influence and control it. The early models of Greek astronomy (Ptolemy, ≈ 300 BC) are an example of modeling activity that attempted to describe the position of planets and stars without developing an accurate understanding of orbital dynamics.

Another historical example is illustrated in Figure 1 that reproduces the regression fit published in the 1929 seminal paper of Edwin Hubble in his attempt to discover fundamental laws of the cosmology that govern the Universe [3]. The straight line is a simple regression fit between the distance and velocity of various stellar objects. This empirical model lead to a revolution in our understanding of celestial dynamics even though the slope shown on the figure was later found to be incorrect by a factor of 10!

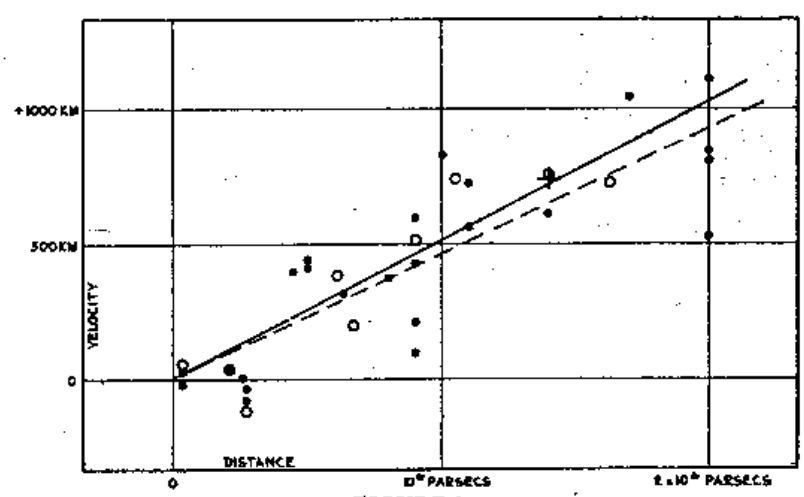


Figure 1. Hubble's empirical model of star velocity versus distance. (From Reference [2].)

The second driver for M&S is to understand the phenomenon observed such that its evolution, assuming, without loss of generality, that this phenomenology is dynamic, can be predicted and controlled. Mechanisms that give rise to the phenomenon must be understood and described with sufficient accuracy, sources of variability must be assessed, and factors that influence and control the phenomenon must be known. Clearly the development of a science-based predictive capability falls in this second category. Predictive models rely on the best-available knowledge of "how something works," not just an empirical description of observations or measurements. Predictive models must also push the boundary of our understanding over multiple time scales, space scales and, possibly, energy scales for reasons that are discussed in section 3.

Figure 2 illustrates a "predictive" capability. Physical measurements of sea surface variability by the European satellite Topex/Poseidon (top half figure) are compared to numerical simulations performed with the POP model at the Los Alamos National Laboratory (bottom half figure). To be capable to predict small variations of sea surface, where the elevation (a few centimeters) is "small" relative to the size of the computational domain (thousands of kilometers), several phenomena must be described and integrated together. These mechanisms include, and this list is not meant to be exhaustive, the chemistry and dynamics of the ocean, atmosphere, and ice caps. Likewise, important couplings include representing the effect of irradiative transfers through the atmosphere and representing the transfers of chemical species, materials, and energy (or temperature) between the sub-models of ice cap and ocean.

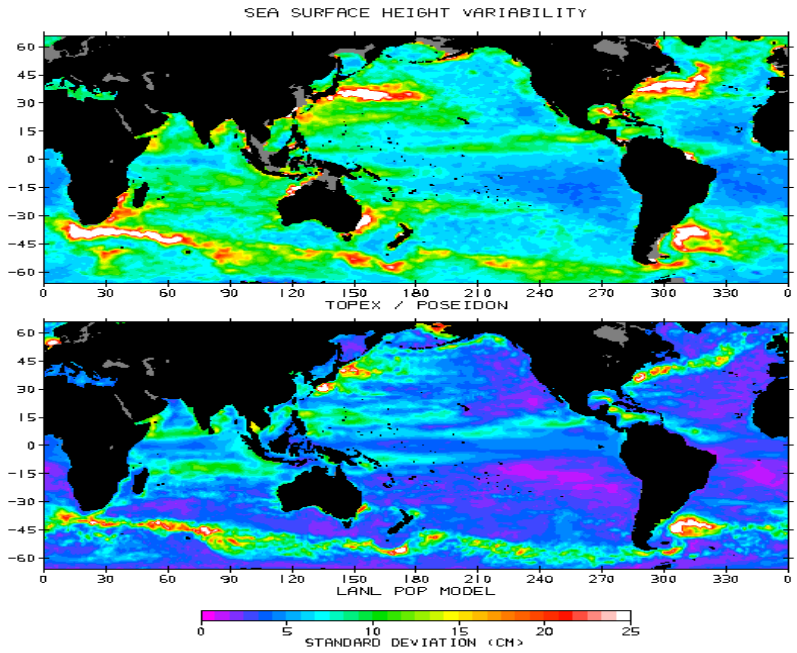


Figure 2. Measurement (top) and prediction (bottom) of sea surface variability.

Clearly, the discussion proposed in this manuscript addresses this second modeling approach where a science-based predictive capability is sought as opposed to the empirical description of a phenomenology. The following bullets discuss, in broad terms, some of the trends currently

observed in attempts to develop science-based predictive capabilities for "hard" sciences such as computational physics and engineering:

- Represent the geometry of the computational domain with the highest possible level of fidelity. Neglect "features" that may not have any significant effect on predictions. Include, however, those that may influence the phenomena being modeled even if this influence only manifests itself at the sub-scale level. (The extent to which "features" must be included, of course, depends on whether feedback mechanisms exist to "link" the various scales over which information gets propagated in the simulation.)

- Implement models and algorithms that describe the behavior of various elements in the problem, based on first-principle physics as opposed to empirical or phenomenological descriptions. In computational physics and engineering, these basic elements include the materials; initial conditions that start the simulation; boundary conditions at edges of the computational domain; energy dissipation mechanisms, whether they are real or artificial; source terms that contribute various forms of energy to, or take away energy from, the problem; and the conservation laws or equations-of-motion that describe the articulation between these various elements.

- Couple the representations of phenomena being modeled in the application. Integrating together several packages may be rendered necessary to account for different effects at play in the problem. The above illustration of coupling models of ocean, atmosphere, and ice caps is an example. It may also be needed to represent various phenomena with their respective, appropriate modeling approaches. For example, one may want to solve the Navier-Stokes equations of fluid dynamics to simulate the aerodynamics of an airfoil while resorting to molecular dynamics simulations to inform on flow properties within the elusive boundary layer nearest to the surface of the airfoil, properties that the conventional equations of fluid dynamics have difficulty resolving. The main difficulty in integrating different phenomena is to avoid "distorting" information that gets transmitted from one model to another.

- Propagate information over multiple and potentially vastly dissimilar length scales, time scales or energy spectra. Bridging the gap between the information content sought at the macroscopic level, where predictions are needed to support decision-making, and information expressed at the sub-scale is necessary to provide closure of the modeling effort. In computational physics and engineering, it would mean bringing closer the continuum (macroscopic) conservation laws and micro-scale models, potentially, all the way to molecular dynamics and inclusion of relativistic or quantum effects. Attempts are currently made to develop algorithmic strategies capable to propagate information back-and-forth, across dissimilar spatial or temporal scales. This is a challenging task because the representation of information appropriate at one given scale may differ from what is appropriate at another scale.

- Propagate uncertainty through numerical simulations to quantify the confidence that decision-makers may place in predictions. Uncertainty is propagated from its identified sources, such as parameters of a model, to predictions. Doing so may require access to formidable computational resources if the uncertainty space to be explored is vast in size and broad in scope. This task is further complicated by the fact that uncertainty may come in many different "flavors," from variability and randomness, to conflicting or non-specific information, to ignorance and lack-of-knowledge. Separating randomness from epistemic uncertainty, understanding which sources of

uncertainty in the problem can be reduced and which cannot, and implementing appropriate techniques for the propagation of each type is, to a great extent, still a matter of open research.

Needless to say that fully implementing these tendencies of science-based "predictability" would require several orders of magnitude more computing power than the resources available for "production" work, that is, what is currently accessed for routine or every-day computing. These requirements are pushing the technology from TeraOPS computing to novel architectures and programming techniques capable to sustain PentaOPS and ExaOPS computing.47 This aspect of "predictability," even though it is critical to the successful deployment of a predictive capability, is no further discussed in this manuscript.[48]

3. The Activities of Resolving, Modeling, and Assumption-making

In this section, we turn our attention to the description of the process that is at the core of any M&S activity and consists in "balancing" the ability to resolve, hopefully, most phenomena of interest in the numerical simulation with the requirement to model some of them. We have seen that the so-called "first-physics" approach to M&S seeks to develop science-based models capable of describing information at the pertinent space scales, time scales, and energy spectra for as many phenomena as reasonable possible. To understand what this principle implies at a conceptual level, numerical simulations can be idealized as simple "filters" that carry information only within specific length, time or energy bandwidths.

[47] TeraOPS, PentaOPS, and ExaOPS are units of computing power that define a peak speed in terms of number of floating point operations (such as a multiplication) sustained every second. They correspond to 10^{+12}, 10^{+15}, and 10^{+18} floating-point operations per second, respectively.

[48] In lieu of a technical discussion, anecdotes are provided, here, to illustrate what this computing power represents. Imagine, first, that everyone in the World could be provided with a pocket calculator and could perform one floating-point multiplication per second, which, as far as humans go, represents computing at a rather fast pace. How long would it take to compute what a one-PentaOPS machine computes every second? Would it take two seconds, two hours or two days? A second anecdote provided for illustration is to estimate how quickly a one-PentaOPS computer could read the 142 Million books (estimated) of the U.S. Library of Congress? Answers are provided at the end of the manuscript.

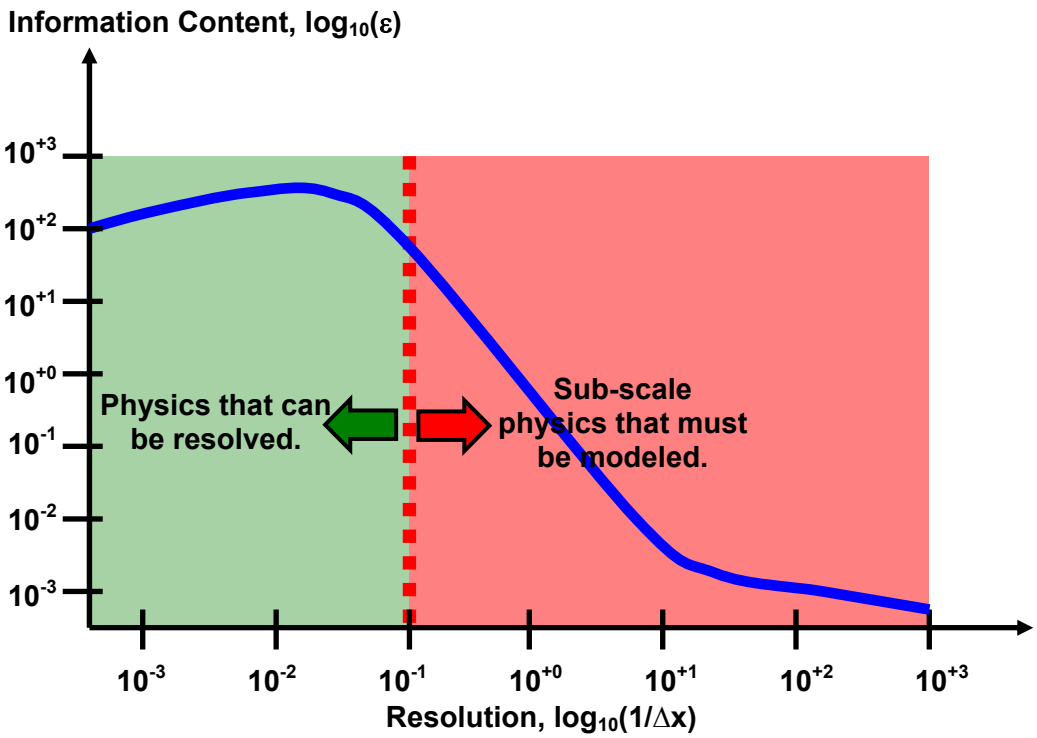


Figure 3. Conceptual view of a simulation as a low-pass *"filter"* of energy or information.

Figure 3 illustrates this concept where the energy or information content that a simulation can capture, labeled "ε" on the vertical axis, is depicted as a notional blue curve as a function of resolution, labeled "1/Δx" on the horizontal axis.49 The figure suggests the existence of a low pass-band region within which information is fully resolved by the computational discretization, followed by another region where the energy content rapidly falls with increasing resolution. It means that numerical simulations are capable to resolve information up to given cut-off, spatial or temporal, frequencies beyond which the phenomena being studied must be modeled.

This naturally leads to the discussion of resolving versus modeling. Performing a numerical simulation capable of resolving all phenomena of interest is clearly not possible. Doing so would ultimately face fundamental barriers and uncertainty principles such as those put forth by Heisenberg in 1927. One must, however, also recognize that information always "leaks out" of macroscopic scales towards the micro-scales, as suggested in Figure 4. To alleviate the lack-of-resolution in numerical simulations, and bring closure to the modeling effort, sub-scale modeling becomes a necessary "evil." Sub-scale models introduce feedback mechanisms such that the fine-granularity energy or information that cannot be resolved on the basis of the computational discretization alone can, instead, be modeled and accounted for in the numerical simulation. In short, sub-scale models stop the "leakage" of information and feed it back to the next level up.

[49] In computational physics and engineering, the symbol "Δx" typically represents a characteristic size of the mesh discretization. By extension it is used, here, to symbolize the level of resolution with which the numerical simulation is carried out whether discretization applies to space, time, energy or any other field.

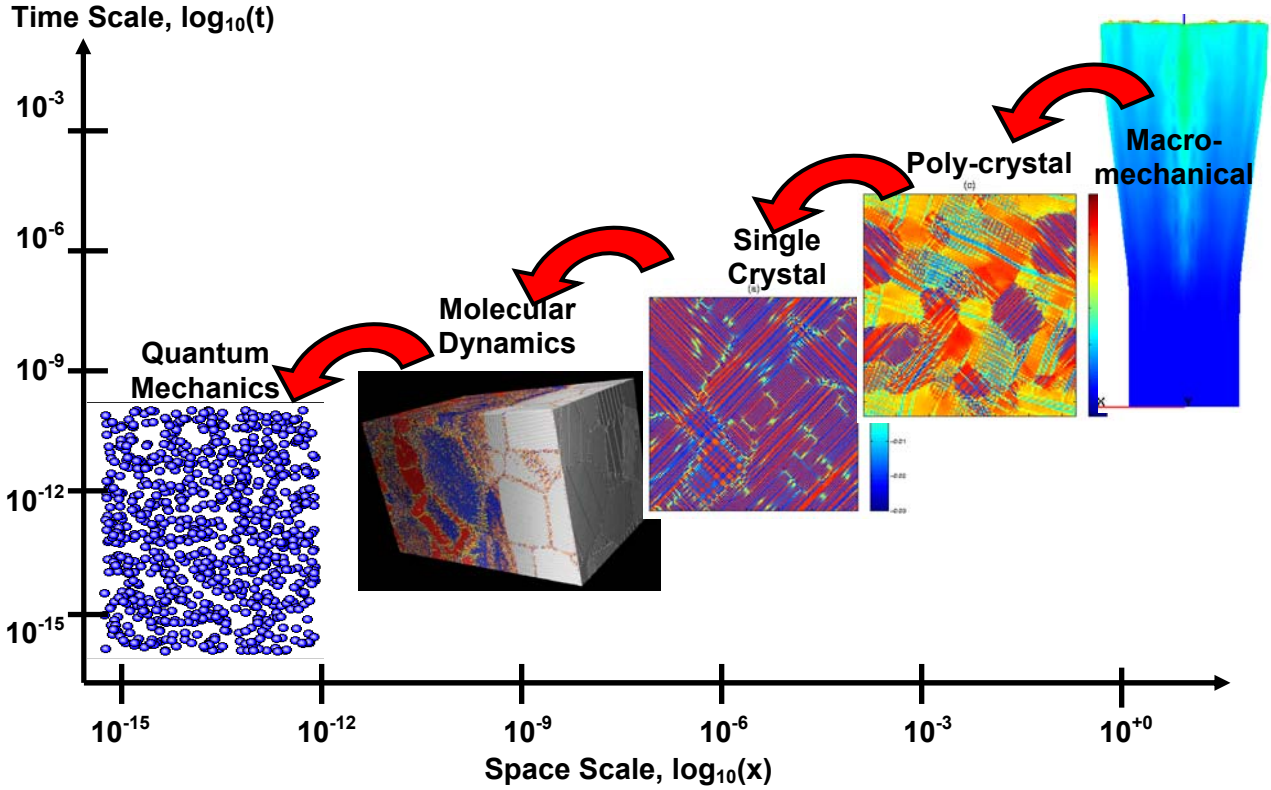
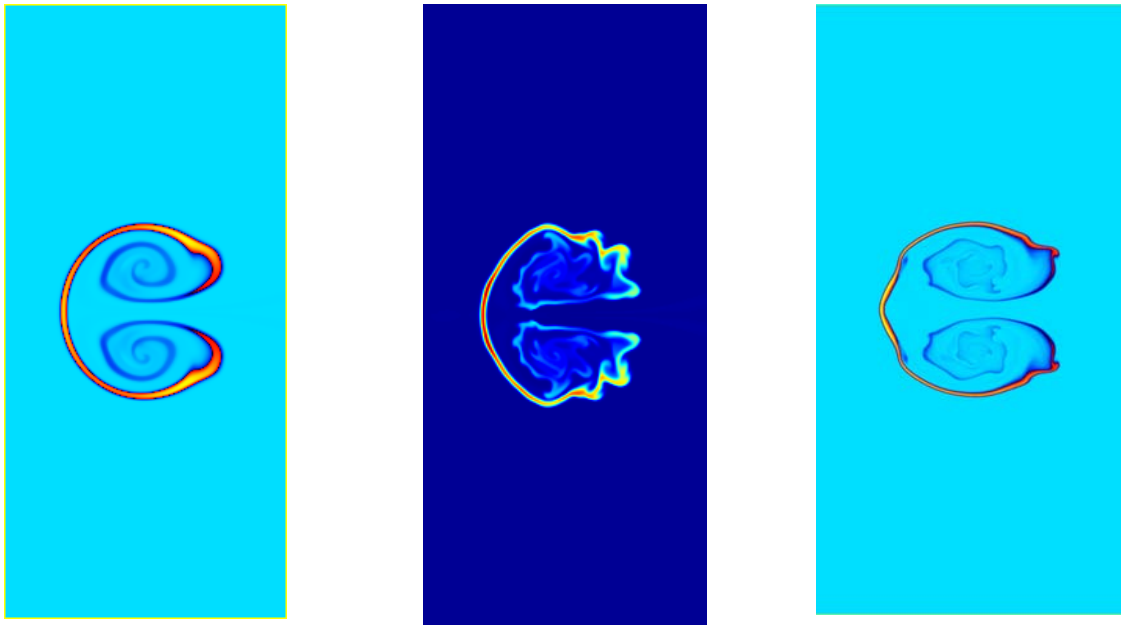


Figure 4. The *"leakage"* of information from the macroscopic scale to the sub-scale, illustrated in the case of numerical simulations of material behavior.

We have established so far that sub-scale models are necessary to provide "closure" of the numerical simulation because information always cascades down from the coarsest-resolution scales to the lower (time, space, energy) scales. But introducing sub-scale models also implies that arbitrary choices and assumptions are made. An illustration of assumption-making is given in Figure 5, where simulations of vorticity in fluid dynamics are compared [4]. These predictions are for the exact same problem. They are, however, obtained using different levels of resolution (125-μm versus 31-μm zoning) and different sub-grid models (piece-wise linear Godunov versus 3rd-order Runge-Kutta interpolation). For these fluid dynamics simulations, the sub-grid models assume the behavior of small elements of fluid located within the computational zone. Hence, they describe a behavior at a scale that the discretization cannot possibly resolve.



(a) 125-μm zoning and linear Godunov model. **(b) 125-μm zoning and 3rd-order Runge-Kutta model.** **(c) 31-μm zoning and linear Godunov model.**

Figure 5. Simulations of fluid vortex illustrating three different combinations of resolution and sub-scale models that result in noticeably different predictions. (From Reference [4].)

The activity of modeling is an exercise in trading-off knowledge for assumptions. In other words, assumptions enable model-building: it is precisely the mechanism by which our ignorance can be masked. Assumptions and arbitrary choices are not just encountered when postulating sub-scale models, as discussed previously. Assumptions are also formulated to simplify a reality of interest and isolate a phenomenon that one wishes to observe through physical measurements or numerical simulations. An illustration is given in Figure 5 that depicts the simplifications and assumptions that the authors made in this 1974 study of the vibration characteristics of the Saturn launcher [5]. The structure is simplified as a one-dimensional beam model with varying cross-sectional areas and moments of inertia depicted, on the figure, with different colors.

It is important to realize that, while unavoidable, modeling assumptions provide us with a false sense of confidence because they tend to "hide" our lack-of-knowledge and the true effect that this ignorance may have on predictions of the numerical simulation. The important question then becomes: "Are our predictions vulnerable to this ignorance?" This is the reason why it is argued that "predictability" should not just be about accuracy, or the ability of predictions to reproduce the available physical observations. It is equally important that predictions be robust to the lack-of-knowledge embodied in our assumptions. Lacking robustness would imply that the predictions vary significantly as one assumption is replaced by another one. Making decisions based on predictions that are potentially sensitive to unwarranted assumptions, in turn, does not inspire great confidence.

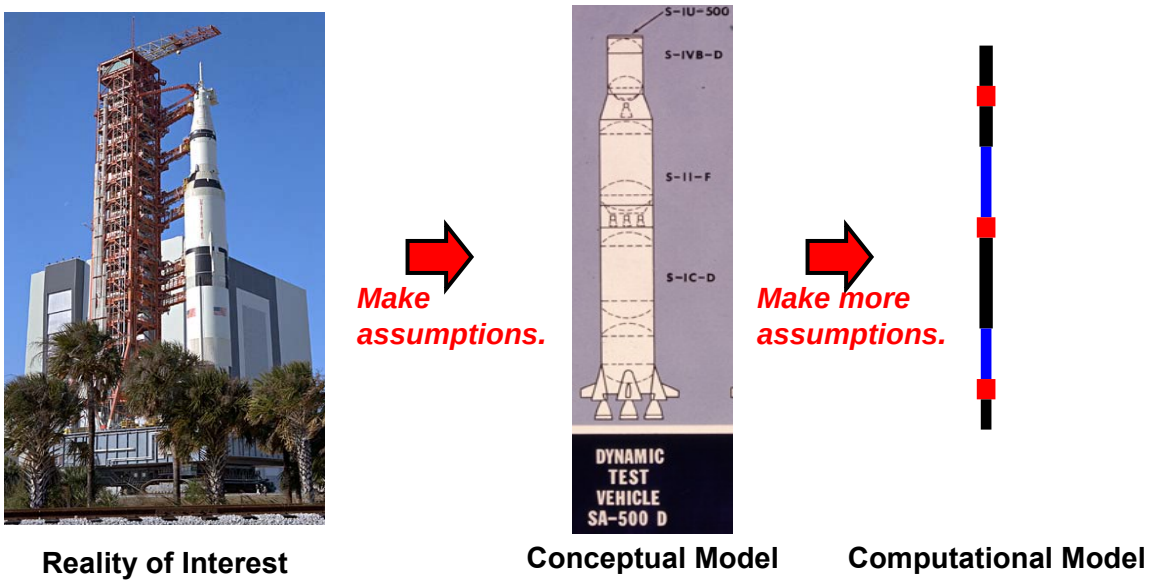


Figure 5. Illustration of assumptions made to develop a computational model and study the vibration characteristics of a launch vehicle. (From Reference [5].)

It is important to realize that, while unavoidable, modeling assumptions provide us with a false sense of confidence because they tend to "hide" our lack-of-knowledge and the true effect that this ignorance may have on predictions of the numerical simulation. The important question then becomes: "Are our predictions vulnerable to this ignorance?" This is the reason why it is argued that "predictability" should not just be about accuracy, or the ability of predictions to reproduce the available physical observations. It is equally important that predictions be robust to the lack-of-knowledge embodied in our assumptions. Lacking robustness would imply that the predictions vary significantly as one assumption is replaced by another one. Making decisions based on predictions that are potentially sensitive to unwarranted assumptions, in turn, does not inspire great confidence.

4. A Brief Discussion of Uncertainty in Modeling and Simulation

In "hard" computational sciences, uncertainty is usually synonymous with randomness. The mathematical theory best accepted, and most commonly encountered, to model, propagate, and quantify uncertainty is probability theory. There are, indeed, good reasons for this. Probability is the most mature of all uncertainty theories, it is supported by the Central Limit theorem [6], and it is backed-up by a formidable amount of experimental evidence. Even though the meaning of a probability may be subject to interpretation, most notably between frequentists and subjectivists (e.g., Bayesian), all interpretations obey the basic same axioms and computational formalism.

Section 4 makes two main points related to Uncertainty Quantification (UQ) and its application to numerical simulations. The first point is to emphasize that, in a context where increasingly more importance is placed on simulating to support decision-making, it is essential to recognize the diversity of uncertainty in our problems. Not everything can be characterized as randomness anymore. Figure 5 illustrates the diversity of uncertainty applied to computational fluid dynamics. Clearly, the three solutions shown are different, even though the exact same equations are solved;

and, surely, this uncertainty is not random in nature because repeating each calculation yields the exact same solution. This uncertainty may be chaotic but it is, for sure, deterministic.

How uncertainty should be modeled mathematically, and dealt with to support decision-making, must account for the origin and nature of information involved. This is because the process of making decisions based on M&S involves combining opinions and expert knowledge; evidence that originates from historical databases, physical observations, and numerical predictions; and, ultimately, consensus-building. Each piece of evidence comes with its own uncertainty, whose nature may differ depending on the type of information involved. Examples include:

- Variability and randomness (naturally occurring, environmental, etc.);
- Multiple possible solutions or observations (also known as non-specificity);
- Ambiguous, vague, or imprecise observations;
- Inconsistencies in knowledge, lack of definitive information;
- *"What is available"* vs. *"what is needed;"*
- Interpolation and extrapolation of models and datasets;
- Inference uncertainty;[50]
- Assumptions made to mitigate the lack-of-knowledge or ignorance; and
- Poorly-known theory and/or first-principle physics.

To support decision-making using numerical simulations, we contend that uncertainty should be defined in a broad sense as ***"everything that is not known absolutely."*** A corollary to the recognition of this diversity is that the theory of probability may not be the only *"tool"* available to model mathematically, propagate, and quantify the uncertainty in our problems. Even though the issue is not debated here, the larger class of mathematical theories of uncertainty known as General Information Theory (GIT) may offer more appropriate representations. The GIT includes such theories such as probability theory, possibility theory, interval arithmetic, fuzzy sets, fuzzy logic, Dempster-Shafer theory of plausibility and belief (or *"evidence"* theory), convex models of uncertainty, and imprecise probability. More thorough descriptions of the GIT are available from References [7-9], to list only a few, and illustrations of this topic can be found in Reference [10].

The second point emphasized in this section is that, to support decision-making, it is essential to treat the uncertainty due to lack-of-knowledge differently from uncertainty that originates from other sources of variability or randomness. We have seen in section 3 that assumptions must be made in numerical simulations to mitigate the lack-of-knowledge or ignorance of some aspects of the problem. Confidently reaching a decision, therefore, necessitates that the predictions be as insensitive as possible, or robust, to these assumptions, as demonstrated in Reference [11]. Otherwise, decision-makers run the risk of basing their decision on information over which they have no control because predictions would change depending on which assumption is used.

We contend that it is a grave mistake to simply ignore the uncertainty introduced in simulations by the modeling assumptions. Unfortunately, it is most often ignored in computational sciences, which

[50] Inference uncertainty means, here, the uncertainty introduced by the entire process implemented to propagate information from what can be observed (such as a measurement) to the targeted (or inferred) quantity. Inference UQ means the propagation of uncertainty from *"inputs"* of the process to its *"outputs."*

leads to a tendency of practitioners to become over-confident in the forecasting power of their simulations. The influence of assumptions on predictions should, instead, be propagated and quantified, as discussed in section 5. One should also be warned against the temptation to treat assumption-making like one would propagate random variability through a problem. Doing so is simply wrong because assumptions are not random variables. The sometimes-suggested approach of "sampling" assumptions with probabilities makes no sense. Attempting, likewise, to aggregate the uncertainty due to assumptions with other sources of uncertainty is meaningless.

The proposal to handle the uncertainty caused by lack-of-knowledge by studying the robustness (or lack thereof) of predictions to modeling assumptions is dealt with in section 5. An example is presented next to illustrate these concepts and setup the discussion. This illustration follows the discussion, in section 3, of resolving versus modeling.

The level of resolution at which a numerical simulation is analyzed defines a solution uncertainty caused by truncation error. To understand its nature and justify why assessing the robustness of predictions is appropriate, the principle of discretization is briefly explained. The essential point is that, when a system of partial (or ordinary) differential equations is solved numerically with a computer code, the solution obtained is different from the solution of the original equations. In fact the solution obtained does not even solve the same equations!

Assume, for example, that a system of one-dimensional conservation laws is solved such as:

$$\boxed{}, \tag{1}$$

where $y^{Exact}(x;t)$ is the exact solution, $F(\bullet)$ denotes the flux term, and $S(x;t)$ is a source or forcing function that drives the dynamics of the system. These generic functions depend on space and time, labeled "x" and "t," respectively. When solving equation (1) with a numerical method, one seeks the best-possible approximation of the exact solution y^{Exact}. This equation applies to a variety of phenomena that include, for example, the Navier-Stokes equations of fluid dynamics, Euler equations of gas dynamics, equations-of-motion of solid mechanics, or the evolution of species in chemical reactions.

The numerical method discretizes the continuous equation (1) on a computational mesh to yield a discrete solution $y_k^n = y(k\cdot\Delta x;\, n\cdot\Delta t)$ where Δx and Δt are the spatial and temporal resolutions, respectively. The approximation y_k^n is obtained by solving a discretized equation that looks, for example, something like:

$$\boxed{}. \tag{2}$$

The approximation shown in equation (2) would be obtained by estimating the operator of time differentiation ($\partial\bullet/\partial t$) with an Euler method, while the differentiation in space ($\partial\bullet/\partial x$) is estimated with a trapezoidal scheme. Many other choices are available in the disciplines of finite volumes, finite differences or finite elements. One may observe that the discretized equation (2) "appears" similar to the original, continuous equation (1). The similarity is, however, misleading.

Contrary to common belief, the discrete solutions y_k^n are not approximations of the continuous solution of y^{Exact} equation (1). Using an analysis technique known as Modified Equation Analysis (MEA), see Reference [12], it can be shown that approximations y_k^n converge to the solution of a modified equation that takes a form such as:

.. (3)

Note that this example is conceptual and the correct form of the modified equation depends on properties of the original equation (1) and numerical method (2) implemented for its resolution. What is important to the discussion is that the solution y of the modified equation (3) is different from y^{Exact}, at least, as long as the spatial and temporal resolutions remain finite ($\Delta x \neq 0$, $\Delta t \neq 0$). The Lax equivalence theorem of Reference [13] details the conditions under which the discrete approximations y_k^n are consistent with, and converge to, the continuous solution y^{Exact}.

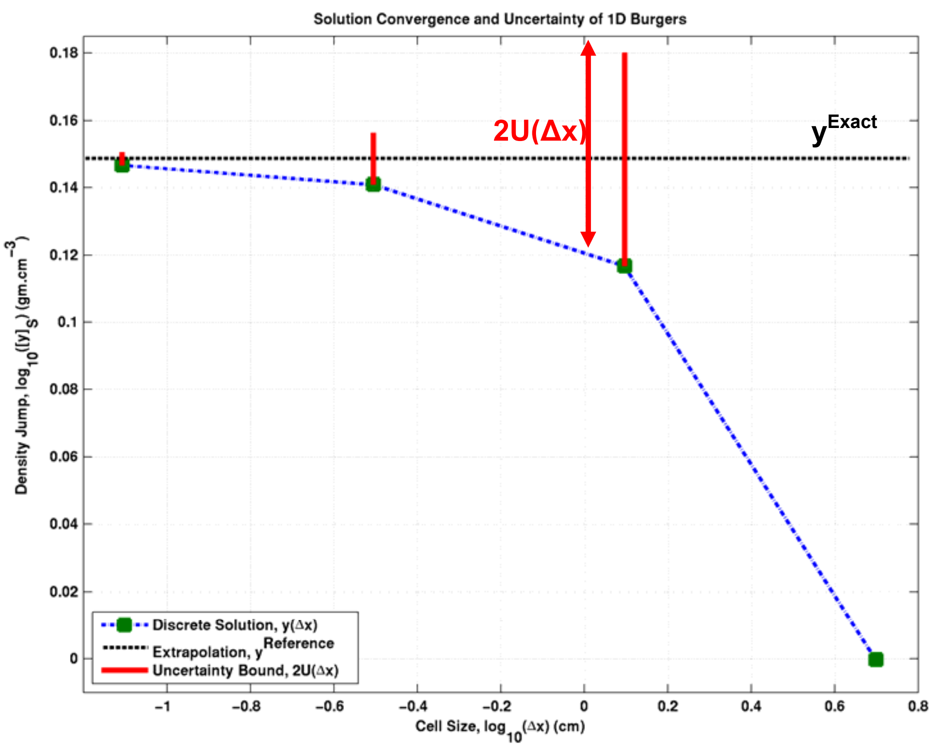


Figure 6. Convergence of discrete solutions y_k^n and bounds of solution uncertainty $U(\Delta x)$ for the resolution of the one-dimensional, non-linear Burgers equation (4).

The terms shown between parentheses in the modified equation (3) represent an infinite series expansion that characterizes the truncation error of the numerical simulation. It can be seen that truncation is what explains the difference between solutions y and y^{Exact}. The effect of truncation is illustrated graphically in Figure 6 that compares four discrete solutions y_k^n, shown with green squares, obtained with different levels of mesh resolution Δx.[51] Because the leading-order term of

[51] For completeness, we note that Figure 6 corresponds to the analysis of a one-dimensional, non-linear Burgers equation where dissipation is added in the right-hand side:

$$\boxed{}\,. \tag{6}$$

In the case of the Burgers equation analyzed (see footnote 51), the exact solution y^{Exact} is known and shown in Figure 6 with a dashed, black line to verify that the bounds of solution uncertainty U(Δx) are accurate.

The application presented briefly in Figure 6 illustrates that settling on a given discretization of the computational domain is an assumption needed to run the computer code, and to mitigate our ignorance of the exact solution. Selecting discretization parameters, such as Δx or Δt, is not an irreducible and aleatoric source of uncertainty. It means, therefore, that there is no such thing as distributions of Δx-values that could be characterized by probability functions, and from which one may draw samples for propagation through the numerical simulations. Likewise, there is no "truth" value of the Δx parameter that could be measured experimentally! To be confident that the numerical simulation provides useful information for decision-making, one should, therefore, establish that the predictions are not influenced by, or sensitive to, this choice of discretization.

To conclude this discussion of uncertainty in M&S, we note that the effect of discretization is to modify the original partial differential equations studied, as can be seen by comparing equations (1) and (3). This is a remarkable observation because it mirrors what has long been known in experimental sciences. Attempting to physically measure a small-scale phenomenon actually changes the reality of interest. As a result, the measurements collected are approximations of what the experimentalists are attempting to observe. Likewise, the discretization of equations on a computational mesh changes the very equation that the numerical analysts are attempting to resolve. As a result, the predictions obtained are only approximations of the true-but-unknown solution of the continuous equations.

5. Trading-off Accuracy, Robustness, and Consistency of Predictions

Section 4 has discussed the diversity of sources of uncertainty in numerical simulations as well as a proposal to treat uncertainty introduced by assumption-making differently from sources of randomness and variability.[52] The rationale is that, in order to be confident in a decision, one should

---

$$\boxed{}\,. \tag{4}$$

The Lax-Wendroff finite volume method is implemented to solve the Burgers equation [14]. The modified equation gives a spatial truncation error equal to:

$$\boxed{}\,, \tag{5}$$

where the contribution to truncation error of time discretization terms is omitted for simplicity. Equation (5) indicates that the Lax-Wendroff numerical method implemented to solve the Burgers equation is second-order accurate because the leading-order term of truncation error is proportional to Δx^2.

[52] Note that this robustness-based treatment of assumptions would be in addition to, not in lieu of, the quantification of sources of uncertainty grounded in randomness and variability. Aleatoric uncertainty can be propagated through numerical simulations using the combination of sensitivity analysis and statistical sampling that, while they may offer practical difficulties when exploring large-dimensional spaces, are very mature technologies in statistical sciences.

guarantee that predictions of the simulation are not influenced by, or sensitive to, specific assumptions. Decisions based on robust predictions are, by definition, more trustworthy than those that may be sensitive to modeling assumptions [11]. This is because robustness mitigates the risk of having selected incorrect assumptions.

In this section, the concepts of accuracy, robustness to lack-of-knowledge, and consistency of predictions are further explored and "linked" together though functional relationships to arrive at a definition of "predictability." The definition proposed is not just a statement; it is backed up by machinery inspired from Reference [11] and based on quantifiable metrics. It is shown how the metrics, that are referred to as the triplet (R; α^*; λ), are used to study the trade-offs between accuracy R, robustness to lack-of-knowledge α^*, and lack-of-consistency of predictions λ.

To start the discussion, we go back to section 2 that evokes "predictability" in computational sciences. We ask what are the benefits expected of a predictive capability. What may be gained by developing a science-based predictive code, as opposed to empirical fits-to-data? A first answer that comes to mind is that science-based simulations lead to improved accuracy of the predictions. Considerable evidence has surely been collected over the past two decades in many disciplines of computational physics and engineering that support this statement. One should be careful, however, that focusing the benefits of M&S solely on accuracy, or the ability of predictions to match the available physical observations, does not open the door to an excessive practice of model calibration or "knob tuning." In fact, it is well-known that too much reliance on matching physical measurements through model calibration is detrimental to generalization, that is, the ability of a numerical simulation to predict new datasets that it has never "seen" before. Even though it improves the goodness-of-fit, calibration is unwelcome if it becomes the means to compensate for unacknowledged modeling errors.

Figure 7. Illustration of modeling strategies, either empirical (left) or science-based (right), developed to calculate the shape of the coastline of Southern Wales.

What else, besides prediction accuracy, or fidelity-to-data, may be gained? We propose, even though this second aspect of "predictability" is somewhat less apparent, that what is also gained is improved robustness of the predictions to modeling assumptions formulated to "hide" the lack-of-knowledge. A science-based, predictive capability tends to produce predictions that are less vulnerable to assumptions made to develop and implement the models and algorithms. To make this point, consider the metaphor illustrated in Figure 7, where mathematical models are formulated to represent the shape of the coastline of Southern Wales. One approach develops an empirical model strongly grounded in physical observations. This is illustrated on the left of Figure 7 with a model that depends on a coarse description of the geometry, through a single radius-of-curvature ζR. The right side of Figure 7 suggests a science-based approach where the representation implements a finer description of the geometry with multiple radius-of-curvature parameters such as $\zeta 1$ and $\zeta 2$.

Irrespective of whether the empirical model (left-side) is more accurate, at least locally, than the science-based model (right-side), the fact remains that its predictions depend on a single, global parameter ζR. A mistake in the knowledge of ζR could have devastating consequences on the accuracy of predicting the coastline's shape. The science-based description, on the other hand, depends on several parameters. Comparable errors in the knowledge of $\zeta 1$ and $\zeta 2$ would have somewhat less detrimental consequences on the accuracy of predictions. It means that, at a similar level of accuracy, the predictions of the science-based simulation are more robust to any potential lack-of-information about how the coastline's geometry should be parameterized.

A third expected benefit of science-based "predictability" is to reach an improved consistency of predictions. Consistency refers, here, to the fact that the predictions obtained with different models, algorithms, or numerical options, do not change significantly. It matters greatly because a consistent body of evidence, including numerical predictions, ultimately leads to confidence in decision-making. Note that, by "confidence," we do not mean the notions defined in statistical sciences of confidence interval or statistical level of confidence. The Oxford dictionary defines confidence as a feeling or belief of certainty. Uncertainty or inconsistency, therefore, reduces confidence. A practice often encountered in sciences and engineering is to solve a problem with different analysis techniques. The argument in favor of such a practice is two-fold. Consistently reaching the same conclusion reduces uncertainty, which increases confidence. Another side of the argument is that, reaching the same conclusion with computer codes or analysis techniques that are based on different assumptions, implies that the final result is robust (or insensitive) to these choices. These relationships can be expressed with mathematical expressions such as:

$$\boxed{} \quad \text{and} \quad \boxed{}, \tag{7}$$

where the symbols λ and α^* denote the lack-of-consistency of predictions and robustness to lack-of-knowledge, respectively. (Definitions for λ and α^* are given below.) Because the sign of the partial derivative is negative, the first of equations (7) expresses compactly that confidence tends to be reduced when λ increases. The second of equations (7) recognizes that confidence tends to increase with α^*. To take advantage of observations formalized in relationships (7), and define a quantifiable process for decision-making, the triplet (R; α^*; λ) is defined next.

The definition of metrics (R; α^*; λ) of fidelity-to-data, robustness, and lack-of-consistency follows closely a discussion given in Reference [16]. It is inspired, to a great extent, by the information-gap

theory for decision-making formulated in Reference [11]. Practical applications of this theory to engineering-like simulations can be found in References [17, 18]. Because of this affiliation, the discussion that follows only summarizes the main points.

Predictions of a numerical simulation are conceptually denoted as $y(p;\theta)$ where the simulation analyzed depends on control parameters "p" and ancillary variables "θ." Variables θ represent the calibration variables, numerical settings, and assumptions that, as we have seen previously, introduce uncertainty in the numerical simulation. This is contrast to the control parameters that define the design space within which the simulation code is exercised. Physical experiments are performed at specified settings of control parameters to collect the observations y^{Test}. With these simple definitions in hand, one can define the metric R of prediction accuracy as:

$$\boxed{}, \tag{8}$$

where $||\bullet||$ denotes a norm, coefficient of correlation, or metric of test-analysis comparison that is deemed appropriate to define accuracy.

The second metric to be defined is the robustness α^*, that is intimately linked to the definition of uncertainty in the problem. For generality, and because our focus is on promoting confidence in predictions as expressed by equations (7), we do not wish to postulate a specific (mathematical) representation of uncertainty such as probabilities, fuzzy sets, or intervals. Instead, families $U\alpha$ of predictive models are defined in a generic sense as:

$$U_\alpha = \left\{ y\left(p;\theta\right) \text{ such that } \left\|\theta\text{-}\theta_o\right\| \le \alpha \right\}, \tag{9}$$

where the subscript $(\bullet)o$ represents the nominal, commonly-accepted, or "best-guess," value of a quantity; and the symbol α is a positive scalar, $\alpha \ge 0$, that represents a horizon-of-uncertainty. The meaning of equation (9) is that the family of predictions $U\alpha$ includes all models or numerical simulations for which the ancillary variables θ do not deviate from the nominal settings θo by more than the horizon-of-uncertainty α. Given additional conditions on the definition of $||\theta–\theta o||$, such as convexity, the definition (9) yields mathematical properties not discussed here but that are essential to the derivation of trade-offs for the triplet (R; α^*; λ). (See References [11, 16].)

The robustness α^* of the numerical simulation, given the definition of uncertainty models $U\alpha$, is the largest horizon-of-uncertainty that can be tolerated while guaranteeing a prediction accuracy no worse than R_{Max}. The significance of the family of models $U\alpha^*$, at the level of robustness α^*, is that any prediction $y(p;\theta)$ chosen within this family is guaranteed to deliver the requirement of prediction accuracy R_{Max}, irrespective of the settings defined for calibration variables, numerical options, and assumptions θ. Robustness α^* is the solution of the nested optimization problem:

$$\alpha^* = \max_{\{\alpha \ge 0\}} \left(\max_{\{\theta,\ \|\theta\text{-}\theta_o\| \le \alpha\}} \left(R \le R_{Max} \text{ such that } y\left(p;\theta\right) \in U_\alpha \right) \right), \tag{10}$$

where the inner-most optimization searches for the worst-possible prediction accuracy that does not, however, exceed the requirement R_{Max}, while the outer-most optimization optimizes the size of the horizon-of-uncertainty parameter α. Contrary to the paradigm of model calibration, the rationale of definition (10) is not to search for settings of the numerical simulation that yield the most accurate predictions. It is, instead, to robust-satisfice prediction accuracy, meaning that one searches for settings θ of the simulation that are as far away from the "best-guess" settings θo as long as

prediction accuracy meets the minimum requirement. One consequence of robust-satisficing is that settings may be found at which predictions of the simulation are more accurate than the prescribed level of error R_{Max} which would be welcomed as a positive windfall.

Once the robustness α* of the numerical simulation has been established, defining the lack-of-consistency λ of predictions is easy. For simplicity, λ can be defined as the range of predictions obtained for all simulations that have the assessed level of robustness, that is:

$$\lambda = \max_{\{y \in U_{\alpha^*}\}}(y) - \min_{\{y \in U_{\alpha^*}\}}(y) \quad (11)$$

A conceptual example is given in Figure 8 to illustrate the fidelity-to-data, robustness, and lack-of-consistency of the triplet (R; α*; λ). The figure assumes that there is a single ancillary variable or assumption θ represented by the horizontal axis. Each blue circle is a prediction made for a particular model, or value of θ. The vertical, dashed line (shown in red) represents the physical observation y^{Test}, from which the maximum prediction error R_{Max} is defined. The figure illustrates that solving for robustness α* is a search for the models, or values of θ, that are located as far away from the nominal setting θo as possible, while yielding predictions that stay within the level of accuracy $|y^{Test} - y(p;\theta)| \leq R_{Max}$. For example, the model that gives to a prediction shown with a black square is outside of the domain of robustness because its accuracy does not meet the requirement The figure also illustrates that the lack-of-consistency λ is given by the range of predictions for all models included in the domain of robustness.

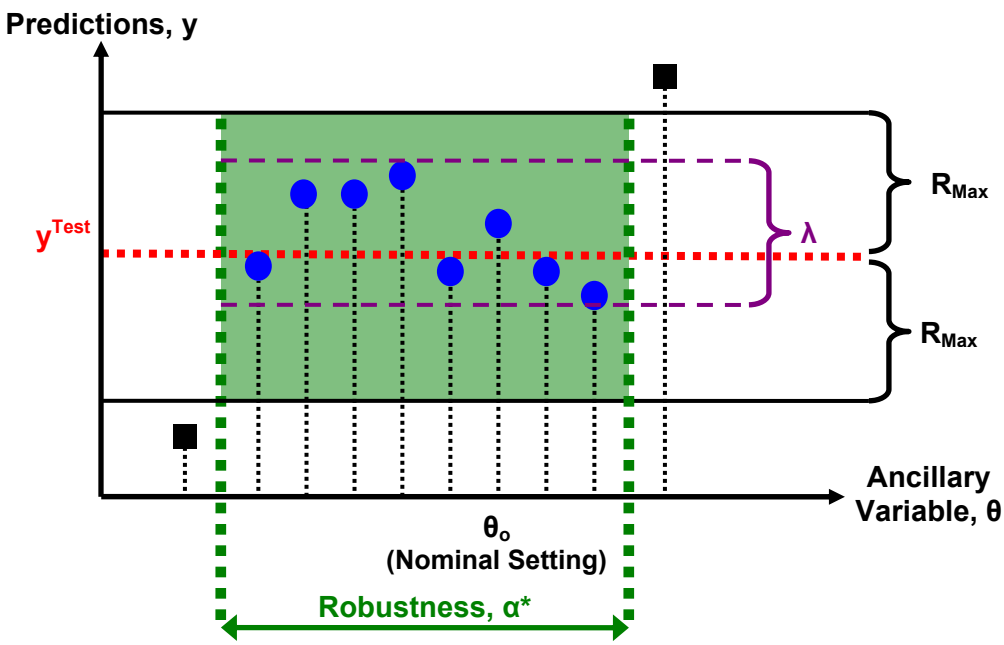


Figure 8. Conceptual illustration of the triplet (R; α*; λ) of fidelity-to-data, robustness, and lack-of-consistency for a one-dimensional source of lack-of-knowledge θ.

Using these definitions for (R; α*; λ), and the information-gap framework [11] to define families Uα of models in equation (9), Reference [16] derives mathematical proofs that explore the trade-offs

between fidelity-to-data R, robustness to lack-of-knowledge α^*, and lack-of-consistency of predictions λ. The main two trade-offs can be summarized compactly as:

$$\boxed{}, \tag{12}$$

and:

$$\boxed{}, \tag{13}$$

to which one may add, by virtue of the chain rule:

$$\boxed{}. \tag{14}$$

These mathematical inequalities lead to the following discussion:

- ***Robustness decreases as fidelity improves.*** Numerical simulations made to better reproduce the available physical observations become more vulnerable to the sources of lack-of-knowledge in the problem that may include potential errors in the modeling assumptions, ignorance of the functional form of models, sub-optimal discretization and resolution settings, and incorrect definition of the numerical options.
- ***Lack-of-consistency increases as robustness improves.*** Numerical simulations or models that are made more immune to the potential sources of lack-of-knowledge in the problem yield less consistent predictions. The inconsistency of predictions would, in turn, reduce confidence in the ability of the family of simulations or models to forecast configurations, settings, or environments that have not been tested experimentally.
- ***Lack-of-consistency decreases as fidelity improves.*** Numerical simulations made to better reproduce the available physical observations lead to more consistent predictions when forecasting configurations that have not been tested experimentally. Although improving the consistency of predictions obtained from a family of equally-robust models is generally a good thing, it could also lead to over-confidence in the ability of the numerical simulation to forecast new scenarios. This trade-off expresses the false sense of confidence provided by an excessive reliance on model calibration.

Equations (12-14) do not indicate that it would be impossible to obtain high fidelity-to-data, high robustness to lack-of-knowledge, and high consistency of predictions. Developing a predictive capability that, eventually, would feature these three desirable attributes of "predictability" may be possible. What these equations do demonstrate, is that simultaneously improving all three attributes is not possible. Our analysis indicates that past physical observations, accompanied by an incomplete understanding of the phenomena measured, cannot unequivocally establish true prediction of the behavior of the system.

We propose a definition of "predictability" grounded in the understanding, and exploration, of the trade-offs between the three attributes (R; α^*; λ). Exploring these trade-offs is essential because the fundamental process by which decisions are made is to balance the search for maximum "reward" with the aversion to "risk" [11]. It is precisely by understanding the trade-offs between reward and risk that humans make decisions. We argue that, similarly, it is by exploring the trade-offs between fidelity-to-data (this would be the "reward" piece) and robustness to lack-of-knowledge (this would be the "risk" piece) that models can be selected for optimal forecasting. The "predictability" of a family of numerical simulations of equal robustness would then be given by the consistency of their predictions (this would be related to metric λ), hence, the definition:

> "Predictability" is the ability to make as-consistent-as-possible predictions from a class of equivalent models, where the "class of equivalency" stems from the fact that all simulations or models included in the class share a minimum requirement of accuracy while being equally robust to the sources of lack-of-knowledge in the problem.

We emphasize that this definition is not just a concept. Quantitative metrics for (R; α^*; λ) and a numerical procedure have been proposed to explore these trade-offs for decision-making. (See Reference [16] for discussion of the theory and References [17, 18] for "real" applications.)

To conclude this section, the procedure outlined above for decision-making is illustrated in the case of a code verification exercise that assesses the numerical performance of, and quantifies the solution uncertainty for, a hydrodynamics code developed at Los Alamos. Figure 9 shows a mesh refinement study conducted for six test problems. These test problems are solved using different levels of mesh resolutions. The prediction errors are obtained by calculating the norms of differences between the (known) exact solutions yExact and numerical predictions y(Δx). The numerical settings varied in these runs include the spatial resolution Δx, temporal resolution Δt, a stability condition, and five other options of the numerical solver. The study performed 12,256 runs over a period of several weeks using up to four processors of a 30-TeraOPS platform. The solid lines of Figure 9 show the evolution of the mean error as the level of spatial resolution is increased (or $\Delta x \rightarrow 0$). The dashed lines represent the ± 1-σ uncertainty bounds that result from varying the time step, stability condition, and other numerical settings of the algorithm. Different colors refer to the analysis of different test problems.

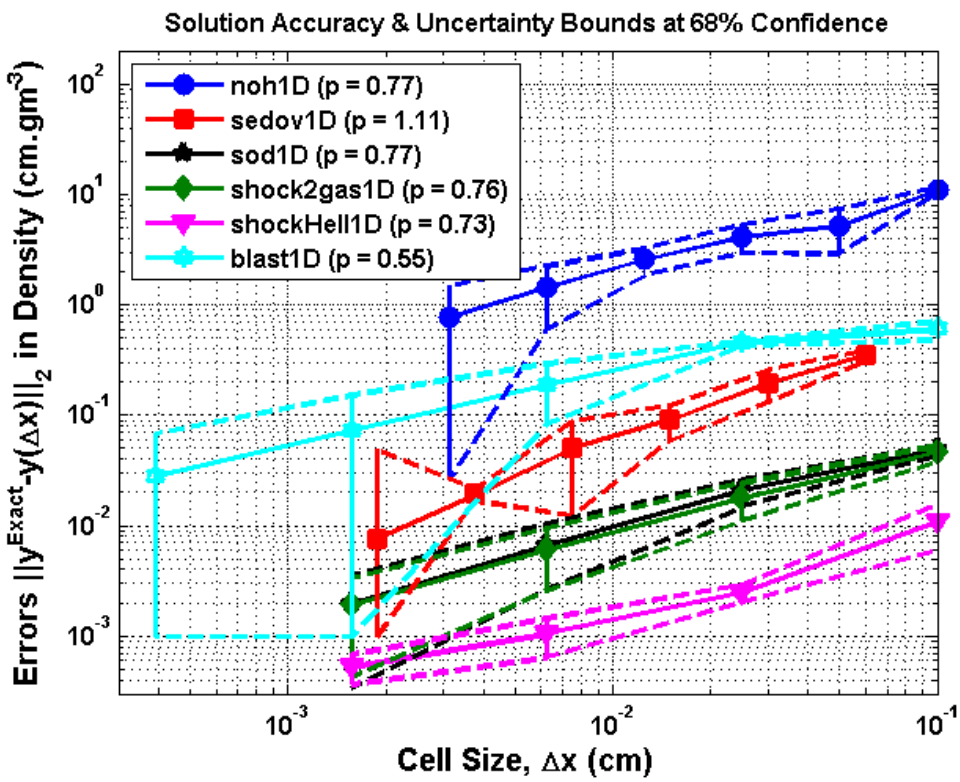


Figure 9. Statistics ($\mu \pm \sigma$) of solution errors versus resolution size Δx.

One immediate observation is that the ± 1-σ bounds of variability of solution accuracy seem to be sensitive to the selection of resolution with which the test problems are analyzed. Running with increased resolution reduces the mean solution error; it also increases the ± 1-σ bounds of uncertainty. Greater accuracy, as $\Delta x \rightarrow 0$, comes at the cost of increased sensitivity to ***how*** the simulation is performed. These results are a manifestation of equation (12) that expresses the trade-off between accuracy and robustness to lack-of-knowledge, where the ignorance applies, here, to the selection of settings of the numerical method.

Data for one of the test problems analyzed, that simulates the propagation of a shock in a perfect gas (labeled "sod1D" in Figure 9), are studied next. For this code verification exercise, the solution accuracy can be defined as $R = ||y^{Exact} - y(\Delta x)||$, where the code predictions are compared to exact solutions of the equations solved, as opposed to physical observations. The most prominent lack-of-knowledge is the level of resolution at which calculations are performed. As argued previously, the discretization size Δx is not a random variable whose uncertainty can be characterized by a probability distribution, nor can it be measured experimentally. The level of resolution that may be most appropriate to perform the numerical simulation is simply unknown. The same observations apply to the other numerical settings varied in this study. To assess the numerical performance of the code, one should therefore understand the extent to which solution accuracy is sensitive, or robust, to these settings.

Figure 10 explores the robustness of solution accuracy to the assumption represented by the choice of resolution size Δx. The solution accuracy on the horizontal axis is plotted as a function of the inverse of resolution, or Δx^{-1}, shown on the vertical axis. Moving up on the vertical axis means performing the calculation with more resolution. On average, increasing resolution tends to reduce

the prediction error, as indicated by the solid, blue line. The green, dashed line and dotted, red line represent, respectively, the best and worst solution errors.

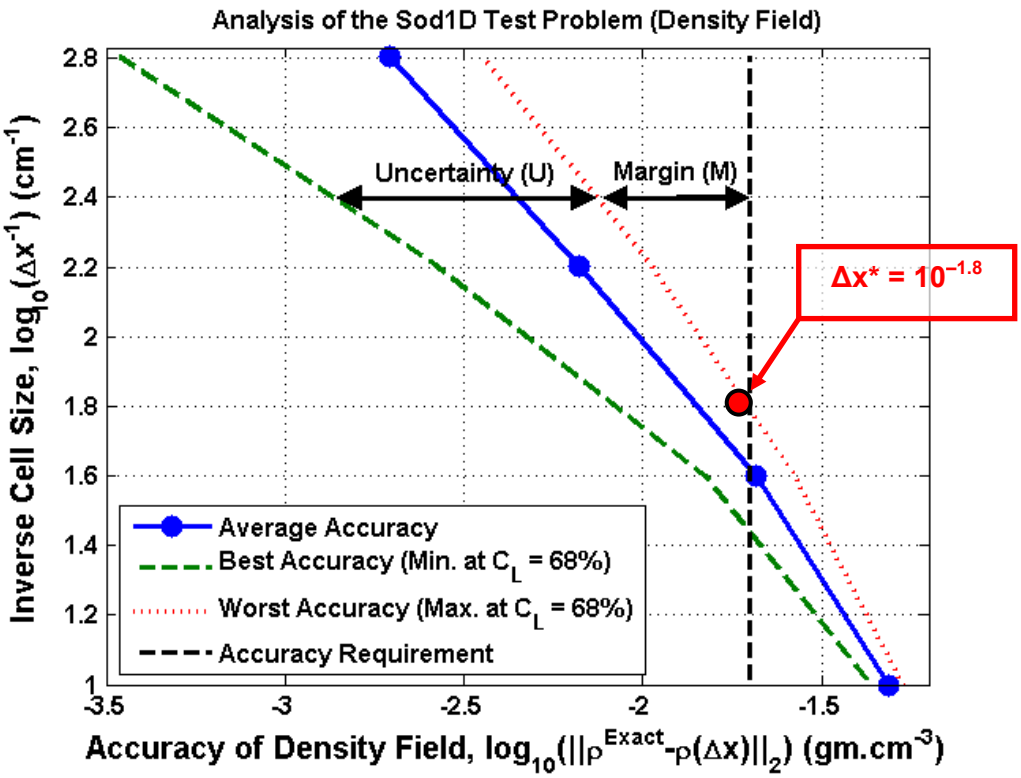


Figure 10. Trade-off between accuracy $||y^{Exact} - y(\Delta x)||$ and robustness to mesh size Δx.

Figure 10 illustrates the protocol proposed for decision-making. A hypothetical requirement of accuracy R_{Max} is shown as a vertical, black line. Given the locus of solution accuracy versus resolution, the analyst can select the robust-optimal size $\log_{10}(\Delta x^{*-1}) = 1.8$, or $\Delta x^* = 0.016$ cm, that ***satisfices*** the requirement of accuracy. The robust-optimal resolution is, by definition, the largest possible value Δx that guarantees the required level of solution accuracy. This point is identified on Figure 10 where the worst-accuracy curve intersects the requirement R_{Max}. It is emphasized that, because this solution is robust-satisficing, it is possible to find other numerical solutions that provide more accurate predictions. These solutions would be located in the range that extends from the worst-case accuracy (shown as a red, dotted curve) to the best-case accuracy (shown as a green, dashed curve), at Δx = constant. This range of predictions from the worst accuracy to the best accuracy is, according to definition (11), equal to the lack-of-consistency λ.

Another advantage of the protocol proposed for decision-making is that it does not give a unique answer. Because it explores the trade-offs of attributes (R; α^*; λ), Figure 10 presents the information necessary to study the effect of selecting an accuracy requirement that differs from the one shown at R_{Max} = 10–1.3 gm.cm–3. A decision-maker must understand this information to select a solution based on his/her desire of "reward" and aversion to "risk." It is by trading-off "reward," or, here, the requirement of minimum accuracy, for "risk," or, here, the combination of robustness and lack-of consistency, that decisions are made.

Answers to the Questions of Section 3

The questions asked in section 3 to illustrate the computing power available in the era of TeraOPS or PentaOPS computing are briefly answered. Beyond this anecdotal evidence, the point made to conclude our discussion is that this formidable power does not suffice to reach "predictability" in computational science. In fact, it may not even be necessary, as the trade-offs between accuracy, robustness to lack-of-knowledge, and consistency of predictions indicate.

The first anecdote asks how long it would take to the World population to compute the same number of floating-point operations that a one-PentaOPS machine can sustain every second. Given an estimate of the World population at 6,697,254,000 (as of September 2010), it would take about 2 full days, or 48 hours, to perform 10+15 operations, assuming that every human can type one operation per second.

To answer the second anecdote (how long would a one-PentaOPS machine take to read the entire U.S. Library of Congress?), one needs to make a few assumptions about the size of this archive. We can reasonably assume 142 Million books with averages of 200 pages-per-book, 450 words-per-page, and 10 letters-per-word. These assumptions are conservative and lead to an upper bound of 1.28 × 10+14 letters to read, approximately. Further assume that one floating-point operation is equivalent to reading one letter or character which, again, seems a conservative estimate. A one-PentaOPS machine would then be able to read this entire collection about 8 times per second!

Acknowledgments

This work is performed under the auspices of the Verification and Validation (V&V) program of Advanced Scientific Computing (ASC) at Los Alamos National Laboratory (LANL). The author is grateful for the continuing support of Mark Anderson, V&V program leader. Leadership and encouragements of Laura McNamara, Sandia National Laboratories, are also acknowledged and greatly appreciated. LANL is operated by the Los Alamos National Security, LLC for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396.

Bibliographical References

[1] Kusnezov, D., Soudah, J., Advanced Simulation and Computing Program Plan Fiscal Year 2005, *Report NA-ASC-101R-04-Vol-1 of the Office of Advanced Simulation and Computing (ASC)*, National Nuclear Security Administration (NNSA) Defense Programs, Washington, D.C., September 2004.

[2] Oden, J.T., Belytschko, T., Fish, J., Hughes, T.J.R., Johnson, C., Keyes, D., Laub, A, Petzold, L., Srolovitz, D., Yip, S., "Simulation-based Engineering Science: Revolutionizing Engineering Science Through Simulation," Report of the *National Science Foundation Blue Ribbon Panel on Simulation-Based Engineering Science*, February 2006.

[3] Hubble, E., "A Relation Between Distance and Radial Velocity Among Extra-galactic Nebulae," *National Academy of Sciences*, Vol. 15, No. 3, March 1929.

[4] Rider, W.J., Greenough, J., "The Accuracy and Efficiency of Shock-capturing Methods for Shocks," *Society of Industrial and Applied Mathematics (SIAM) Computational Sciences and Engineering Meeting*, San Diego, California, 2002. (LA-UR-03-0778.)

[5] Collins, J.D., Hart, G.C., Hasselman, T.K., Kennedy, B., "Statistical Identification of Structures," *AIAA Journal*, Vol. 12, No. 2, February 1974, pp. 185-190.

[6] Laplace, P.S., "Mémoire sur la Probabilité des Causes par les Evènements," *Mémoires de l'Académie des Sciences de Paris VI*, 1774, pp. 621-656.

[7] Joslyn, C.A., Booker, J.M., "Generalized Information Theory for Engineering Modeling and Simulation," Chapter 9, *Engineering Design Reliability Handbook*, CRC Press, 2004.

[8] Singpurwalla, N.D., Booker, J.M., "Membership Functions and Probability of Fuzzy Sets," *Journal of the American Statistical Association*, Vol. 99, 2004, pp. 867-889.

[9] Ross, T.J., Booker, J.M., Hemez, F.M., Anderson, M.C., Reardon, B.J., Joslyn, C.A., "Quantifying Total Uncertainty Using Different Mathematical Theories in a Validation Assessment," *9th ASCE Joint Specialty Conference on Probabilistic Mechanics and Structural Reliability*, Albuquerque, New Mexico, July 26-28, 2004. (LA-UR-04-2862.)

[10] Hemez, F.M., "Answering the Question of Sufficiency: How Much Uncertainty is Enough?," *1st International Conference on Uncertainty in Structural Dynamics*, University of Sheffield, United Kingdom, June 11-13, 2007. (LA-UR-07-3575.)

[11] Ben-Haim, Y., Information-gap Decision Theory, Academic Press, Elsevier Publishers, 2nd edition, 2006.

[12] Hirt, C.W., Shannon, J.P., "Free-surface Stress Conditions for Incompressible-flow Calculations," *Journal of Computational Physics*, Vol. 2, 1968, pp. 403-411.

[13] Lax, P.D., Richtmyer, R.D., "Survey of the Stability of Linear Finite Difference Equations," *Communications in Pure and Applied Mathematics*, Vol. 9, 1956, pp. 267-293.

[14] Lax, P.D, Wendroff, B., "Systems of Conservation Laws," *Communications in Pure and Applied Mathematics*, Vol. 13, 1960, pp. 217-237.

[15] Hemez, F.M., "Numerical Uncertainty in Computational Engineering and Physics," *2nd International Conference on Uncertainty in Structural Dynamics*, Sheffield, United Kingdom, June 15-17, 2009. (LA-UR-09-1964.)

[16] Hemez, F.M., "The Myth of Science-based Predictive Modeling," *Foundations 2004 Workshop for Verification, Validation, and Accreditation (VV&A) in the 21st Century*, Arizona State University, Tempe, Arizona, October 13-15, 2004. (LA-UR-04-6829.)

[17] Hemez, F.M., Ben-Haim, Y., "Info-gap Robustness for the Correlation of Tests and Simulations of a Non-linear Transient," *Mechanical Systems and Signal Processing*, Vol. 18, March 2004, pp. 1443-1467. (LA-UR-02-3538.)

[18] Hemez, F.M., "A Decision-centric Analysis of the Sandia Structural Mechanics Validation Problem," *48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and 9th AIAA Non-deterministic Approaches Conference*, Waikiki (Oahu), Hawai'i, April 23-26, 2007. (LA-UR-07-2385.)

Phillip Huxtable, Leveraging Computational Social Science for National Security[53]

Presented at the DTRA/Sandia Workshop on Challenges in Computational Social Science. Santa Fe, New Mexico. October 25-28, 2010.

Introduction

McNamara and Trucano's[54] discussion of the challenges of incorporating computational social science methods into national security decision-making identifies many crucial issues and suggests potential approaches to resolving them. There is very little in their paper with which I categorically disagree. In this paper, rather, I offer my perspectives on many of the issues that they raise, and offer alternative interpretations of their character and potential solutions. I whole-heartedly agree with their stance of "productive skepticism" and fully believe that more effective decision-making will result from robust debate on the best applications of social science methods (computational and others) to the significant security challenges of today.

In particular I will argue below that in the national security decision-making domain, the differences between social science and physical system modeling and between academic and national security applications of computational modeling have significant implications for best leveraging these approaches. I will make suggestions for 1) how academia could adapt to smooth the transition of concepts and approaches into government decision-making process; 2) how government-funded R&D projects should be structured; 3) how analytic organizations should recruit, train, and deploy its workforce; and 4) how verification and validation should be conducted.

---

[53] The positions and opinions stated in this paper are solely those of the author and do not represent the positions of the United States Government or the U.S. Department of Defense.
[54] References to McNamara and Trucano are to *DTRA Challenges Report* 8/4/10 Draft.

Before developing my arguments, several disclaimers are in order. My experience is rather limited, and primarily with smaller Department of Defense (DoD) organizations that provide analytic support to primarily operationally-focused components. My principal focus has been on effective policy implementation, not policy-making. Those from organizations with other missions and analytic demands, larger budgets and/or larger, more highly educated staffs may well have different perspectives. Thus, the opinions stated below are intended primarily to apply to analysis in support of military operations and not specifically to other applications of computational social science such as strategy and policy-making, mission rehearsal and training. In my opinion, however, the issues raised may be relevant in other contexts as well. In addition, while McNamara and Trucano (6) differentiate computational social science from quantitative social science, here I don't make that distinction. I will suggest below that expertise in techniques from formal modeling, statistical analysis, system dynamics, agent-based modeling and other approaches are necessary for an analysis team to tackle the national security questions it will be called upon to address. Finally, while I will speak in terms of broad generalizations, I recognize that many individuals and groups long ago moved past my recommendations and are conducting excellent work. My purpose here is to generate discussion about the best path forward, not to criticize nor denigrate anyone's work.

Computational Social Science and National Security

As McNamara and Trucano (20) point out, "computational modeling is part of defense culture" and most decision-makers are comfortable incorporating insights from mathematical models into their decision calculus. On the other hand, there are a considerable number of national security officials (including both civilians and military officers) who have little trust in,

or use for, mathematical models of human behavior. Indeed, it often seems that officials fall into two factions: those who are not skeptical enough, and those who dismiss all quantitative social science as either invalid or unnecessary. The challenge of this paper, and of this workshop, is to find the balance between the two camps that takes maximum benefit from the methods of social science.

Military organizations have been employing mathematical models for diverse purposes since long before computers were invented, but my impression is that it wasn't until the 1960s that quantitative techniques were applied by the U.S. military to analysis of social phenomena. It is also my impression that the benefits of quantitative social science were exaggerated and they under-delivered during the Vietnam era, creating a backlash against this approach in the post-war period, although some social scientific approaches continued in areas like analysis of U.S. forces and nuclear deterrence.

It is clear that quantitative social science is again gaining favor among the defense establishment, for a number of reasons. First, as McNamara and Trucano point out, decision-makers are looking for more scientific (i.e. unbiased and objective) sources of information on which to base their decisions (20). As long as social scientists don't exaggerate their ability to discern objective truth, social science can indeed contribute to a better foundation for decision-making. Second, enormous gains in computing power have made more complex (and more realistic) models of social phenomena possible, although this has the perverse effect of making the understanding of insights derived from the models more difficult. Third, advances in information exploitation and information management technologies allows researchers and analysts easy access to types of data that were formerly much more difficult to obtain.

While this is a propitious time for quantitative social science, proponents should be careful not to repeat the mistakes of the post-Vietnam era. Workshops like this that seek to develop high standards and hold computational social science accountable to meeting them are very important for establishing the proper place of these approaches in the array of tools available to support national security decision-making.

There is potentially great value to be gained from deeper incorporation of computational social science, and quantitative social science more generally into analysis in support of decision-making processes. First, quantitative methods have a long history of uncovering patterns in the data derived from observations of complex phenomena, and these techniques can be quite useful for gaining insight about new or poorly understood systems. It is the use of social science's hypothesis testing methods, however, that I believe is the most powerful use of computational methods. It is true, as McNamara and Trucano state (29) that "models and simulations…are not arbiters of truth in any absolute sense," but perhaps they can be leveraged to uncover the falsehoods inherent in faulty qualitative system characterizations offered in support of decision-making.

In general, the purposes behind the application of computational social science approaches in academic settings are different than the needs of national security decision-makers. The academic social sciences tend to favor individual research in narrow domains over collaborative, multi-discipline research during promotion and tenure evaluations while national security decision-makers require holistic analysis of their problem domains. It seems to me that cross-department collaboration tends to involve the computer science department with a single social scientist. Secondly, most social science sub-disciplines, especially those that favor computational and quantitative approaches, tend to favor research agenda that focus on

developing grand theory and finding broadly-based empirical support over deep understanding of a single case while analysis supporting national security focuses on the single cases almost exclusively.

Thirdly, as McNamara and Trucano point out, the fields like geophysics, seismology, and physics that have most rigorously addressed issues of modeling and simulation have strong applied subfields (25). In the social sciences however, the applied sub-fields tend to be more focused on qualitative approaches, and to use quantitative and computational methods mostly for policy evaluation. In addition, social science has not paid sufficient attention to problems of missing or suspect information. Quantitative and computational social scientists tend to choose research domains where the data is already collected, or fairly easily collected while national security analysts are often dealing with situations with great uncertainty derived from both missing and unreliable information.

Finally, and perhaps most importantly, academic social science is primarily concerned with understanding "what is" not with projecting "what if." While accepting that the act of observing an event may affect the outcome, social scientists are generally focused primarily on understanding a situation or phenomenon *as it exists* and perhaps how it will evolve into the future but expend very little effort on understanding how various interventions might change the outcome. This however is precisely the task of the national security analyst: to understand how interventions (including by allies, neutral, or competitors) might change a situation into one more or less favorable to the analyst's country. This of course raises issues concerning the proper role of academic institutions, and for the professional codes of ethics of various disciplines that I will not address here. Nonetheless, there are implications for graduate student training and research

agenda that could be considered without causing harm to academic integrity or field researcher safety.

While McNamara and Trucano point out many of the parallels between modeling physical systems—both natural and human-made (e.g. weather, earthquakes, nuclear weapons)—and social systems, the differences are significant for understanding how best to incorporate the insights gained from years of experience with modeling those physical systems. The most significant difference is the absence of immutable laws in the social domain. Put simply, people make choices and they learn, in ways that don't apply to physical systems.[55] Two groups of people don't necessarily behave in the same manner even though the situations may appear identical. Of course I won't deny that there are patterns in human behavior but, and this is especially significant for analysis of national security issues, people learn from experiences and change their behavior to gain an advantage in future interactions.

I believe that these differences between modeling physical systems (which national security officials are comfortable with) and social systems imply needed modifications of the way we approach application of computational social science. While I'll discuss in more detail in a later section, here I'll offer my opinion that the principal implication is that (expensive) large-scale, complex, multi-system models may not be useful beyond a narrow scope. As Robert Axtell *et al* noted (quoted in McNamara and Trucano 36) social science models are usually created *de novo* for each research project. While they suggest this derives from the creative impulse in the modeler, I would argue that the primary driver is that each situation is unique in significant ways. Social science helps identify the factors that should be considered in the model building activity, but the driving factors and their arrangement will likely vary significantly, even when the situation looks quite similar to that encompassed in an extant model.

[55] I'll leave out discussions of animal- and machine-learning as irrelevant distractions from the main points here.

Suggestions for Improving Applications of Computational Social Science

In this section I will suggest modifications to the manner that academic departments, U.S. Government R&D agencies, and analytic organizations conduct business in order to better take advantage of the potential of computational social science. As I stated above, I clearly recognize that many organizations have already taken steps in the right direction, and many have conceived and implemented ideas superior to those offered here. I see much reason to be optimistic that the trends indicate that the potential benefits of computational modeling will be more effectively realized in the near future.

Based on shifts in the types of candidates interested in positions in national security organizations my sense is that graduate students and (more slowly) their academic departments are moving toward greater acceptance of national security careers, perhaps because of declining opportunities for new PhDs in academia but also because of an increased enthusiasm for public service. As noted above, universities could better prepare their students for analytic careers in national security by developing and exercising their abilities to model and evaluate multiple potential futures. Furthermore, by rewarding multi-discipline, multi-contributor research agenda with promotions and tenure departments could incentivize their faculty to produce research much needed by the national security establishment. I believe that we have an immense body of theory to explain individual social phenomena, but insufficient development of methods that integrate the insights of multiple disciplines into a coherent, actionable whole.

It is my impression that considerably more multi-discipline research occurs in universities today than in previous periods, in part due perhaps to the lure of government research funding that increasingly demands such approaches. While this has been quite

beneficial to the advancement of computational social science, I believe that modifications to their approaches, some of which are being applied at some agencies, would have significant payoffs.

First: Methods before software tools. Some large projects funded by the US government have included multiple social science disciplines and approaches (a good thing) and have attempted to develop tools to integrate the insights from the various methods. While the resulting systems have been impressive (from a computer science perspective) at handling data and supplying the outputs of some models as inputs of others, it is not clear that the systems are all that useful for supporting national security decision-making. As McNamara and Trucano assert multiple times, model outputs have to be interpreted to be useful for decision support. Using model outputs as inputs to other models concerns me for two reasons: because uncertainty and error is compounded mathematically in unknown ways and because the difficulty of interpreting the final model result is raised considerably. It has been my experience that the most useful social science software was often developed by the social scientist himself/herself to solve a particular problem.[56] My suggestion is that funding agencies push for development of analytic processes that integrate insights from multiple disciplines to support decision-making in specific domains, first conceptually, followed by development of mathematical (or other computational) representations, before moving to development of robust, user-friendly tools. Rigorous peer review and validation should occur at each step.

Second: Flexibility is important. As noted above, while there are important similarities among cases it's likely that each case is different in significant ways, and those differences should be represented in models used to analyze the case. Analysts shouldn't have to write new

[56] Even though the code was likely very ugly and didn't conform to any standards. This describes some of my early work pretty well.

code from the bottom up to represent their current problem, but nor should the government spend hundreds of thousands of dollars (or more) to develop a system that is only valid for a narrow domain. I suggest government support for the development of systems that facilitate model development, analysis, and visualization in a large number of contexts and that support development of libraries of models and components that can be shared and understood easily among analysts.

Third: The "over the fence model" doesn't work. McNamara and Trucano (42) note the difficulties of transitioning technologies developed in isolation from the end-user. Everyone in the community knows this approach doesn't work, yet the vast majority of analytic capability projects are executed this way, and most are unlikely to transition in any way that makes use of their apparent potential. The U.S. Defense Department has recognized this problem however and DoD agencies are being pressured to increase attention to transition. There are three ways to overcome the over the fence problem (but two are unlikely to be fully successful).

Government organizations with analytic missions could opt to develop new capabilities internally, involving their production analysts in the R&D processes. This is unlikely to be successful, however. The demand for analysis always exceeds an organization's capacity, so the temptation is always to delay R&D activities to complete production tasks. Furthermore, the motivation and skills necessary to conceive of and implement innovate new analytic techniques are more likely to be found in universities and commercial businesses than within government agencies.

A second solution to this problem is to contract with the organization that develops a capability to supply the workforce needed to conduct analysis. This approach has worked in the past, with private companies and semi-governmental organizations providing much needed

analytic expertise and tools to support national security decision-making. Given current trends toward less out-sourcing and more limited defense budgets, this solution is probably not optimal in the long term.

A final approach to transition is to continue to centralize capability development around the government R&D organizations, but to involve end-user organizations earlier and more deeply in the development and evaluation process. This requires sustained effort on both government organizations as well as the contractors but pays off in that end-users get their requirements met earlier in the process, allowing earlier, smoother, and more effective transition. Training time is shortened because the target organization's initial adopters are learning tools and techniques during system evaluation and validation activities. This is crucial for acceptance by the end user analysts. I have seen many sophisticated capabilities rejected by analysts because they had no basis to trust the tool or method's validity and usefulness for their analytic tasks. Analysts who work with a capability throughout its development cycle gain an understanding of its strengths and weaknesses and thus develop confidence in using it for tasks on which significant decisions (and their professional reputations) will be based.

Government organizations could also adapt to better utilize computational social science in their analytic processes, and indeed many have begun. Much has been made of the Defense Department and U.S. Intelligence Community's attempts to recruit and develop analysts with deep cultural knowledge. It seems to me, however, that the demand for social scientists with computational and quantitative skills is increasing as well. The problems and decisions with which organizations are required to grapple are increasing complex and multi-faceted. I would suggest that the analytic teams assigned to support these decisions be multi-faceted as well, including members with a variety of analytic skills, social science disciplines, and relevant

cultural knowledge. People tend to be most comfortable working with people with similar educational backgrounds but just as I suggested that academics be incentivized to develop methods of integrating multi-discipline research, analytic organizations would do well to develop multi-faceted approaches to analysis as well.

The increased use of computational and quantitative social science methods would have implications for the recruitment and professional development of the government workforce as well. Some analysis processes will simply require highly educated, experienced analysts to execute properly; others will not. I would argue that we should not build tools and methods aimed at the workforce we have; we should strive to develop the most powerful analytic techniques possible, demonstrate their value, and then decide whether the increased costs are acceptable.

Finally I will turn to the subject of Verification and Validation with which McNamara and Trucano are much concerned. I agree with them that verification is quite important. One should be diligent in ensuring that a tool is operating as one expects (although I have had a few experiences in which significant errors in software code had only minor effects on analytic results). And I agree that validation is also important, though I don't think it should get quite the attention that they do.

Building on a central theme of McNamara and Trucano's paper, the contribution of analysis to decision-making is a social process. Within that process are several validity checks. Decision-makers are seldom concerned with the sophistication of the analytic methods behind a recommendation; they are more concerned with the persuasiveness of the logic of the recommendation in natural language. Secondly, analytic organizations and the analysts within them live on their reputations; one shoddy analysis can wound an organization or a career.

Analysts, and their supervisors, have strong incentives to produce quality analysis. In my opinion validity would be better supported through better educated and experienced analysts than through formal external validation.

As I suggested above, each case is likely to require substantively different models than similar cases; the interpretations that can validly be drawn from a model will vary according to the specifics of the case. I am skeptical that many computational models will be developed and used in multiple analysis tasks. I am not saying that validity should not be a driving concern for developers and analysts. Rather I am suggesting that the payoffs from formal external validation be carefully considered against the costs.

Concluding Remarks

As discussed above, computational social science has the potential to significantly improve the analysis supporting national security decision-making. Properly implemented, and accompanied by the appropriate caveats and disclaimers, these approaches begin to fill many of the gaps in our analytic capabilities. While we should always strive for and demand excellence, we should also remember that the threats facing us are enormous, and coming at us quickly. We can't afford to wait for nearly perfect solutions many years from now but should be prepared to field partial solutions now as long as we are confident we understand their limitations and can assess those effects on the certainty of our recommendations.

Jeffrey C. Johnson, Grave Error: Towards a Science of Error in Social Network Analysis and Modeling in the National Security Context

Presented at the DTRA/Sandia Workshop on Challenges in Computational Social Science. Santa Fe, New Mexico. October 25-28, 2010.

In most of basic scientific research errors in modeling are problematic, but not life threatening. In fact, the risks associated with pushing forth the envelope of scientific knowledge are rife with risk, uncertainty and errors of various kinds. But this tolerance of error declines precipitously as we move from basic science to applied science. Although based on earlier more theoretical work, applications of science should ensure that bridges won't fail, aircraft not drop out of the sky, and electronic consumer items won't cause cancer. Thus, as we move into the world of modeling of humans we need to be sure that the scientific theory surrounding the phenomena of interest is well enough developed to give confidence that outcomes can be determined with some degree of certainty and reflect some reasonable reality. Furthermore, errors of various kinds are inevitable, but the science of errors should be well understood so that more reasonable estimates can be made. In statistics, for example, the behavior of error under random sampling is well understood and can be used to make estimates about population parameters from a sample taken from the population, as long as the sample is truly random.

In the national security context errors can be costly. Innocent lives lost, unjust incarcerations, wasted resources, and alienation of various populations are among the potential costs. Thus, if we use models of human behavior we should have a through understanding of their reliability and validity and the reliability and validity of the data that feeds them so as to not have a "structural" failure as in the bridge reference above. These models should transition from the more theoretical world of risk and uncertainty to an applied world of known model performance and a fundamental understanding of errors that can undermine the validity of such models. This is

particularly the case for social network models that are often used for targeting individuals and shaping social terrain in the course of fighting terrorism, insurgencies, etc. What this paper addresses is a call for the science of errors in social network models in the national security context where errors can have grave implications. Just as an understanding of error is essential in proper conduct of classical probability and statistics, error in all of its aspects should be understood in the conduct of social network modeling and analysis.

Social Network Modeling as Computational Modeling

Given this collection of papers is concerned with computational models in the social sciences the question arises is social network analysis and modeling a kind of computational approach? The National Academy of Science committee on "Modeling from Individuals to Societies" was tasked with reviewing the state of the art in modeling of humans at the macro, meso, and micro levels. In the course of discussing social and behavioral models it was clear that committee members used different terms and focused on different elements of modeling leading to some confusion. To focus the discussion and organize future discussions (and the eventual book), data was elicited from committee members to understand the behavioral modeling domain. Members were asked to name all the types of models they could think of. The top 36 most frequently mentioned models were then used for a pile sort task. Each member was asked to sort the 36 types of models into piles according to how similar they perceived the models to be to one another (using an online data collection program). Data in this form was amenable to Multidimensional Scaling (MDS). An MDS of the similarity among types of models revealed three primary clusters, computational models, mathematical models and what was termed cultural models. Within the computational cluster included agent based modeling, but also dynamic social network analysis and simply social

network models (See Figure 1). So for the purposes of this paper we consider social networks models and analysis to fall within the computational framework.

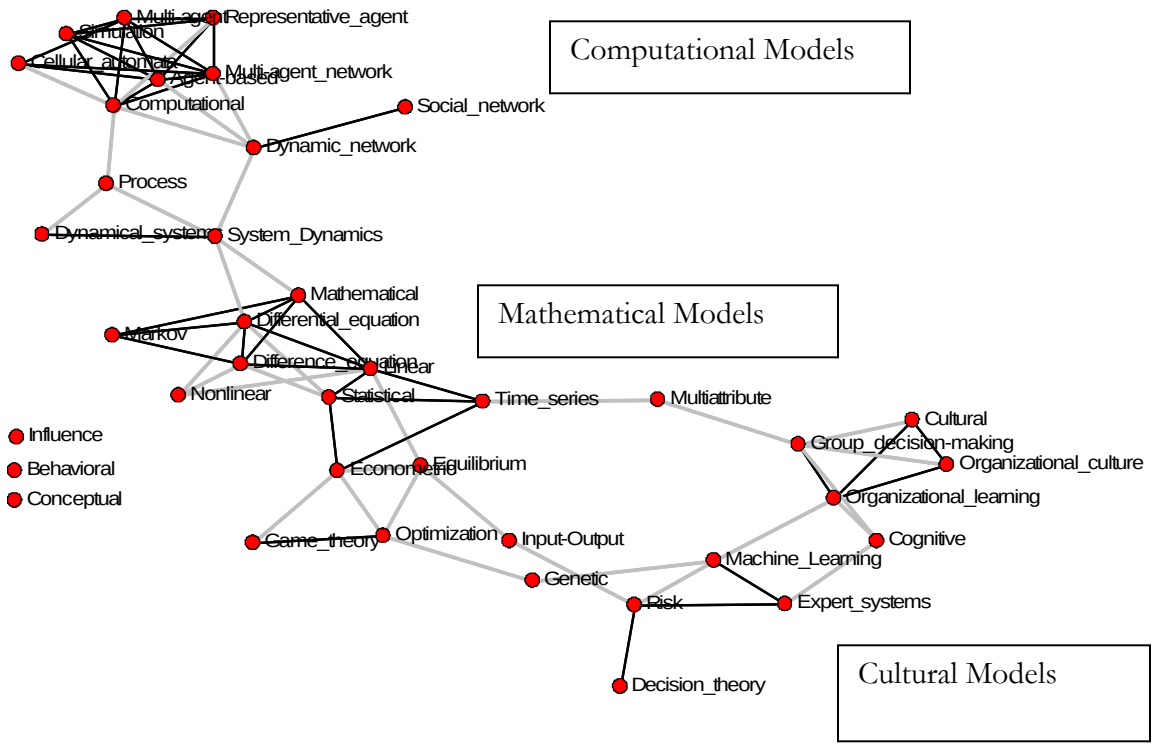


Figure 1. MDS of the similarity among types of models as perceived by NAS committee members.

The Potential for Error

Errors in network data can arise from a multitude of sources. The way questions were framed, the manner in which network boundaries are specified, the willingness of respondents to answer questions, the manner in which data is aggregated, informant accuracy, the erroneous attribution of behaviors, to name a few, can all impact error in terms of missing data or even the presence of data that lacks validity and may be misleading. These problems are even more pronounced in the national security context where data can come from a variety of sources all of varying reliability and the databases in which they are housed can be a mix of various measures at various levels. As an

article in Defense Update, an online defense magazine published in Israel, on targeting in networks for counter IED (http://defense-update.com/newscast/0308/news/news0703_iednetworks.htm) stated:

> In the modern and developed world, where most of those support networks operate, government agencies and the business sector generate unprecedented volumes of data. Customer profiles, organizational operational performance, and personal behavior of individuals are monitored by multiple service providers. Some data resides in structured form in databases or exists as real-time streams. Some exists in unstructured form, for example as e-mails, electronic documents or media files. Whatever the form, there exists huge potential to transform these data into relevant intelligence to improve business decision-making.
>
> Before the solution could be deployed, JIEDDO had to address various data challenges. Until recently, much of the data received from multiple sources in theater, from DOD and other U.S. government agencies, was not integrated or coordinated with data developed by other units or agencies, it usually remained unstructured or lacked a common format or vocabulary. Also, data quality was problematic because of the amount that was manually keyed or handwritten, the lack of standard format and templates, and the variety of sources.

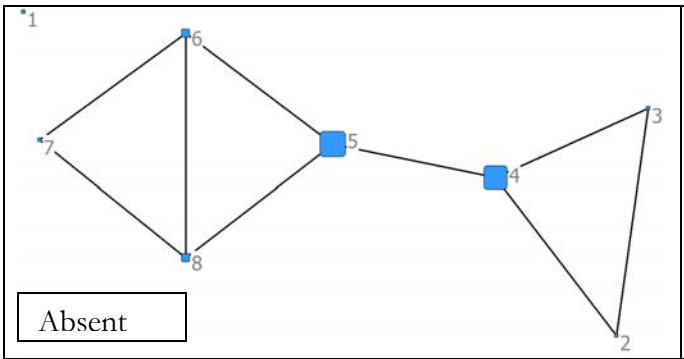
One problem with data that is critical for the validity of computational based network models is the inclusion of edges or links in network models when in fact they do not exist or can be mistakenly attributed to an ego (focal node) based on the behavior of other nodes in the network. This is particularly important in the national security context in that various sources of data (e.g., SIGINT, HUMINT) are used to establish the existence of a link in a model or analysis. So, for example, Signal Intelligence (SIGINT) relies on observed linkages among electronic devices while Human Intelligence (HUMINT) relies on various human based sources for establishing the presence or absence of linkages among individuals. Within each of these various domains there are various issues of data reliability and validity, but when combined a whole slew of other issues arise, particularly the assignment of devices to human actors in the network. What is problematic is that analysis or model results of these networks are used to make life and death decisions about the targeting of individuals.

Towards a Science of Error in Social Networks

Social network measures vary in their robustness in the face of error (Borgatti, Carley and Krackhardt, 2006). A simple measure like degree centrality tends to be relatively robust while some path based measures such as betweenness centrality are more vulnerable to error. This is important in light of the fact that betweenness centrality reflects control and brokerage roles of an individual, a very important measure for identifying key nodes, possibly for targeting. Figure 2 shows an example of the sensitivity of betweenness to the presence or absence of a single bridging tie or edge. A missing tie between nodes 4 and 5 in the top figure would completely hide the brokering importance of these two nodes. Conversely an erroneous commission of a tie between nodes between 3 and 8 can lead to node 3 have more importance then it actually does in the network (e.g., attributing communication behaviors to 3 that are actually those of node 4).

In an attempt to understand the impact of these omissions and commission errors on various network measures, Johnson, Boster and Holbert (1989) used a Monte Carlo simulation approach to study error in networks derived from snowball samples employing a fixed choice methodology. An important finding of this study was that degree centrality was relatively robust under different sampling conditions. Borgatti, Carley and Krackhardt (2006) have looked at the issue of the error afforded by omission and commission errors of both edges and nodes in the collection of network data. They found that errors in various centrality measures resulting from random exclusion and inclusion of edges in random graphs varies as a function of characteristics of the network itself (e.g., density, sparseness) and accuracy of measures declines predictably with the amount of error introduced. Although an important contribution to the problem it fails to directly address some of the issues raised here in two fundamental ways. First, the implication of these errors needs to be more thoroughly explored in graphs of various types. How these errors behave in

random graphs may be very different than how they behave in small world or scale free networks. Second, the focus of the errors has been at the network or macro level. This includes things like density or aspects of the overall distribution of errors across a particular kind of centrality like, for example, betweenness centrality in a regression model. There needs to be more of a focus on the error at the nodal or micro level to better understand the characteristics and implications of error in a targeting context. It is at the nodal level that targeting decisions are made so it is at the nodal level that error should be better understood and modeled.



	Present	Absent
1	0	0
2	0	0
3	0.15873	0
4	0.095238	0.380952
5	0.126984	0.428571
6	0.039683	0.095238
7	0	0
8	0.246032	0.095238

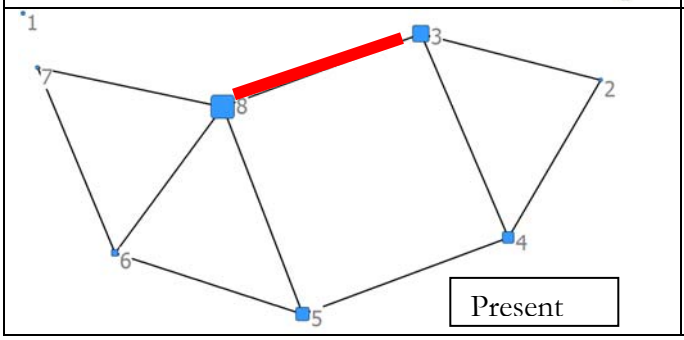


Figure 2. Graphs showing the effect of an added tie on the importance of node 3 with the simple addition of a single edge with respect to betweenness centrality (above).

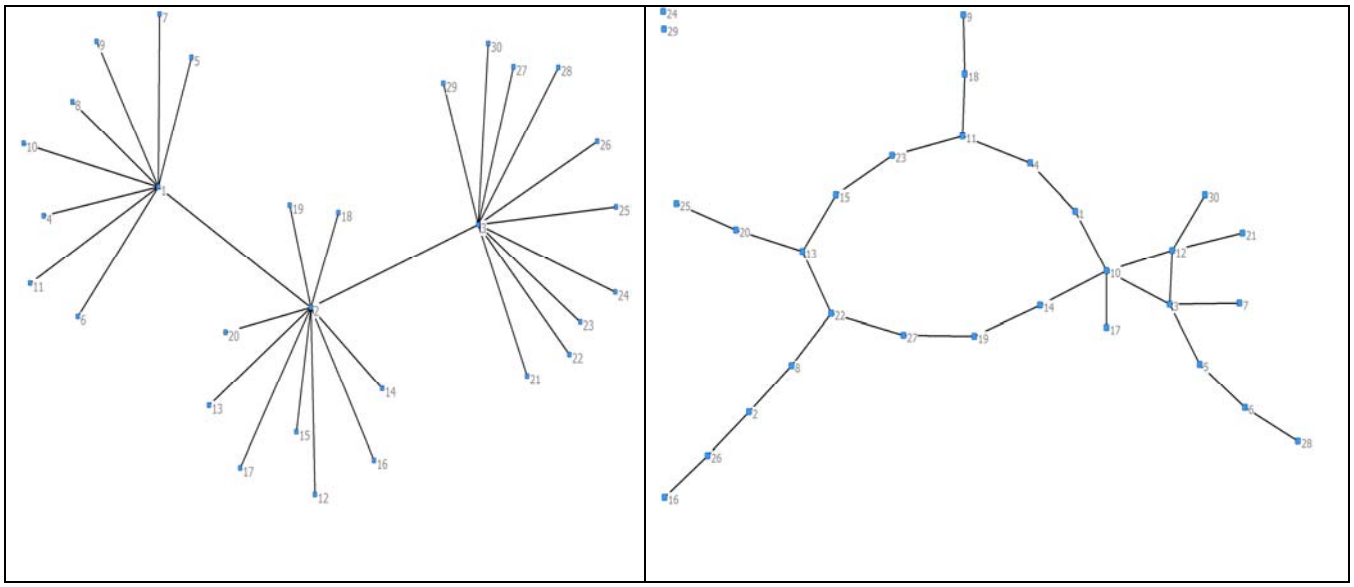


Figure 3. An example of a network derived from a data mining program (SAS in http://defense-update.com/newscast/0308/news/news0703_iednetworks.htm) and an Erdos-Renyi random graph with the same number of nodes and density. The graphs have very different structures that have implications for the impact of data errors on analysis and modeling.

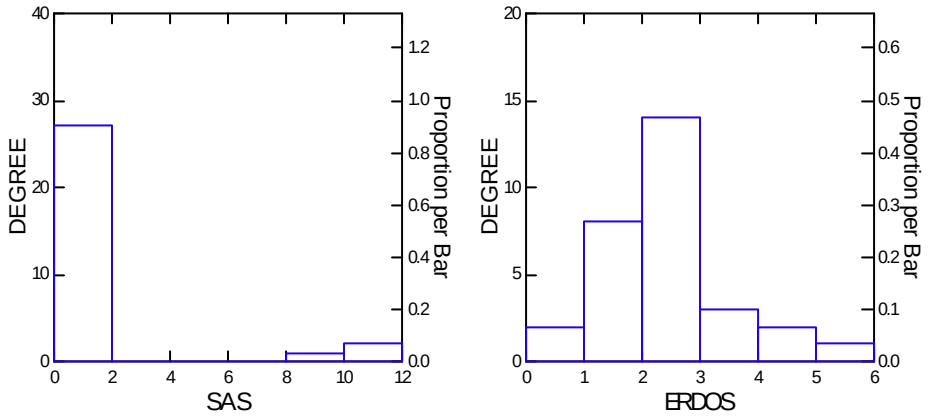


Figure 4. Degree distributions for the SAS text example and a comparable Erdos-Renyi random graph.

Figures 3 and 4 show the implications of the how the differences in network structures may influence how error could impact any analysis. Figure 4 shows two networks one derived from a datamining example (http://defenseupdate.com/newscast/0308/news/news0703_iednetworks.htm)

and the other an Erdos-Renyi random graph for a network of the same size and density as the former but links assigned at random. Note that the two have wildly different structures with the data mining example have an extremely skewed degree distribution (something that might be considered scale free if in a larger network) while the random graph has a degree distribution that is more normally distributed. As such missing edges or nodes would have a much greater impact in the first network then in the chain like structure of the second.

Compared to the collection of other types of data in the social sciences (e.g., attribute based survey data) the collection of social network data can be quite challenging. A major threat to validity in social network research stems from problems of missing data that are due to a number of different sources at a number of different stages in the research process. In addition, errors can arise from data (e.g., a network tie) that is erroneously included, or commission errors. These sources of error all can lead to model misspecification or worse yet, the identification and targeting of actors in a network that are not important or critical to the function of that network, thereby wasting valuable resources or creating backfire effects (the targeting of innocent individuals turns hearts and minds).

One major contributor to missing data is non-response in network surveys (although this can be even more of a problem in data derived from interrogations). Missing data can enter into the picture if the network boundaries are not properly specified on theoretical or other grounds. Network surveys are extremely susceptible to non-response bias in that missing actors and their links can affect structural and analytical outcomes at both the network and individual levels. Respondents can refuse participation, can refuse to answer some or all network survey questions due to such things as interviewee burden or question sensitivity and may drop out of a longitudinal study prematurely as a result.

The design of the study and subsequent sample or instrument design (e.g., types and forms of relational questions) for a given social network problem and context can also be important in limiting threats to validity (and this can vary cross-culturally). Issues of respondent reliability and accuracy have clearly been shown to produce error of various kinds (but the error is often well behaved as discussed below).

We need to be aware of factors that minimize threats to validity in the collection of social network data, particularly in the complete network context and particularly in the security context. This becomes even more important as we use more methods of automated elicitation such as web scraping and other types of data mining. Network data can come from a variety of different sources but it generally boils down to a distinction between primary versus secondary types of data. Secondary sources are those that already exist somewhere in print (e.g., fish exchange records, historical marriage records) or can be found electronically (e.g., Enron emails, Social Networking pages, newspaper articles) using human based or machine based collection. Secondary data by its historical and/or fixed nature dictates and limits the type of relations and levels of measurement that can be used in the course of the research. Primary data collection allows a greater deal of flexibility in the type, measurement and number of relations to be studied.

The following paragraphs provide a discussion of some types of errors that are of concern. This is by no means an exhaustive list, but it does begin to address some of the more important types of errors found in data used in social network analysis and modeling, particularly in the national security context.

Boundary specification errors: When doing whole or complete network analysis there must be inclusion and exclusion rules so that social networks can be bounded for analysis. In more theoretical work, the network is bounded by theoretical focus or the problem at hand. So, the study of small group dynamics at a polar research station might bound the network to include only those

actors that will winter-over together. This doesn't mean there are not other alters in an ego's network (e.g., family that an actor emails, summer workers), it is just that for the purposes of the study the winter-overs are what constitutes the group. This boundary specification could be a threat to the validity of the study if the other types of relations not included in the study are actually important in accounting for variability in the dependent variable. But the bounding of social networks in the national security context is not normally driven by theoretical concerns, but more practical concerns or may even be arbitrary. It is important to note that these decisions will have an effect on both the presence and absence of edges and nodes in any analysis.

Omission errors: Missing edges and nodes can have huge impact on errors in network assessments, particularly for some centrality measures used in the selection of nodes for targeting. These missing data can make networks appear to be more disconnected than they really are or make other nodes and edges in the network appear to be more important than they really are (as evidenced by the missing of a single tie between nodes 4 and 5 in Figure 2).

Commission errors: Like omission errors the erroneous inclusion of nodes and edges can effect the ultimate determination of node level measures and the identification of key nodes (as is clear in Figure 2).

Edge/node attribution errors: Assigning a behavior or attributing something to either an edge or node in a network. Miss assignment of a behavior to a node can yield attributed linkages in a network that in reality do not exist. Although not central to the discussion on network models, attribution error is a common problem in the use of link analysis. The mixing of what we call 1-mode and 2-mode data can be problematic and lead to misattribution.

Multiplex errors: Social networks are multiplex. That is an individual actor's social network consists of family ties, friendship ties, work ties, recreation ties, coffee drinking ties, etc.

Caution should be taken in combining these ties or using one type of relation to infer something about the presence or absence of another type of relation.

Data collection and retrospective errors: Caution should be taken when using network data collected from individuals where the network elicitation question deals with reports of behavior, particularly on social interactions of a discreet nature. So, for example, questions that of the kind "who are the people you interacted with yesterday in the plaza?" are notoriously prone to error. Bernard, Killworth and Sailer (1977, 1980,1982 and Killworth and Bernard (1976 and 1979) (here collectively referred to as BKS) conducted a series of network studies on informant accuracy in social networks involving fraternity members, ham radio operators, and deaf people communicating with teletype machines to mention a few. They basically found that people were inaccurate in their reporting of interactions with others. Thus, ham radio operators, who kept logs of radio conversations, reported both omission and commission errors in their retrospective reporting of radio interactions. BKS asked the operators to list all the people they talked to on the radio yesterday and they could check the accuracy of the reported communications with the actual communications.

This research led to a fury of other research on the topic looking at the relationship between reports of network interactions and accuracy. An important study by Freeman, Romney and Freeman (1987) and Romney and Freeman (1987) found that informants are more accurate in reporting long term patterns of behavior rather than discreet behaviors at some point in time. They noted the participants in a colloquia series at University of California Irvine throughout the quarter. On the day after the last colloquium of the quarter the people who attended were asked to list all the participants present at the colloquium the day before. There were inaccuracies as expected, but these inaccuracies were patterned and predictable. Omission errors were primarily people who normally don't attend the colloquium but happened to be at the last one, while commission errors

were primarily people who usually come to the colloquium but were there for the final one. Thus, individual informants were reporting more on what usually happens rather than on what happened during a specific colloquium. So, the question that was asked at the beginning of this section should be more like "who are the people you usually interact with in the plaza?" or "who are the people you interacted most with in the plaza over the last two weeks?" These reports of long-term patterns of behavior are much less prone to error.

The implications for this in the national security context are simple. Awareness of the way linkages between and among individuals is critical for assessing their potential accuracy. If data was collected from individual's interactions involving specific or well delineated time points, such self reports should be treated cautiously and triangulated against other sources of data. The more recent research on ego biases in cognitive networks (Kumbasar, Romney and Batchelder (1994); Krackhardt (1987, 1990); Johnson and Orbach (2002)) has shown that some individuals in the network are more accurate about reporting linkages than others. Active, more powerful nodes tend to be more accurate. These all have implications on methods for assessing and weighting the reliability and validity of network data and for fixing missing data problems (this doesn't even begin to address issues of deception in social network data).

Data management/data entry: Errors due to data entry and transcription/ translation are well known in other analytical and modeling domains. These can be even more problematic in the network context. These types of errors are well known but important nonetheless.

Data fusion/aggregation: Decisions often have to be made on aggregating data at different temporal, relational and spatial scales. Such aggregations, if done improperly, can create errors at a variety of levels. For example, when aggregating longitudinal real time or streaming data for analysis important individual nodes and edges may be excluded or nodes and edges of lesser importance included that have lost their importance in the network. As in the boundary

specification problem, there should be some guiding principles, preferably of a theoretical nature, for making aggregation decisions.

Error multiplier effects: Errors can interact in ways to make any analytical outcome impossible to determine. Thus, errors can propagate through the data making it impossible to know error impacts or even how one might go about dealing with error (e.g., model error).

Error in secondary sources and data mining: Various forms of secondary source data have inherent biases and these potential biases should be considered in any analysis. Secondary source data collection can be easier (data mining), but it can be fraught with errors at a variety of levels. Examples of important questions include: Do dyadic ties in records have the same meaning (e.g., emails)?; Are nodes really the same? (Phones used in communication, although accurately reflect device to device dyads, may not be the same actors from one communication to the next); Does the observance of two individuals at same event infer a tie?; Are records temporally comparable, at the same scale, etc.?

Often records actually document non-events (e.g., Congressional Record) or represent a reconstruction of the past or an event to meet some agenda making a group or a single actor look good in light of poor outcomes which may include scapegoating and false attribution). Thus, records may be biased in that they are constructed to fit some agenda or reflect actor biases (e.g., newspaper reports). For example, Johnson in his work at the South Pole on small group dynamics reviewed manger's end of year reports and found them to be riddled with inaccuracies. It was understandable in that these reports were actually an attempt to put a more positive spin on the winter events to make the manager look good and to place blame for any problems on others.

Formatting errors: In data mining or web scraping efforts there are errors that can be due to differences in document or web site formatting that can lead to the over or under representation of terms, actors, attributes, etc. in the data retrieval process. Care should be taken that any relations

assigned among nodes is not an artifact of formatting errors. In addition, web scraping and automated data mining methods should be scrutinized for consistency in the operationalization of important concepts. The bottom line is model quality is a function of data quality; garbage in garbage out.

Potential Solutions

There needs to be a serious effort to understand the varying effects of errors in social network data on network model performance. The work of Borgatti, Carley and Krackhardt (2006) and Johnson, Boster and Holbert (1989) should be expanded to look at how errors of various kinds behave across a family of network structures. This line of analysis can include a further expansion of simulation efforts and at the more micro level of analysis various forms of sensitivity analysis to examine how errors of various types impact the robustness of measures for specific nodes. This is a particularly important approach in that the robustness of network measures of targeted nodes can be determined to establish more confidence in any targeting decisions. In addition, work on reliability analysis and more specifically on determining consensus and in estimating the correct answer from patterns of responses (or from different sources) may yield better estimates of various network measures under uncertainty (Romney, Weller, Batchelder 1986).

Simulations: The types of simulations conducted by Borgatti, Carley, Krackghardt (2006) and Johnson, Boster and Holbert (1989) should be expanded and extended to a range of errors and types of network structures. Johnson, Boster and Holbert (1989), for example, found that errors associated with differences in snowball sampling parameters could be modeled as quadratic functions in both hierarchical and non-hierarchical network structures. In this case errors in sampling could be mathematically determined in predictable ways. Similarly, Borgatti, Carley and Krackhardt (2006) were able to statistically model the random removal and placement of edges in

random graphs suggesting that confidence intervals could then be established for various estimates of network centralities. This line of analysis needs to be expanded to include the effect of these kinds of omission and commission errors on other families of network structures such as small world networks, core-periphery networks, scale free networks, etc.

Sensitivity Analysis: Since targeting occurs at the individual node or micro level sensitivity analysis using a simulation approach may be more appropriate. In this form of analysis the robustness of individual nodes with respect to various network measures can be determined though simulations in which links are randomly added and removed. In the course of these repeated removals if there is little change in the network measures of interest then confidence in the structural importance of that node can be determined. Again, this may involve the construction of confidence intervals around the network estimate for a given node. Although not a panacea by any means, it begins at least to systematically address potential impact of errors on targeting decisions. However, there still needs to be caution in that the propagation of errors could make this a fool's errand given the already deep problems with the data. This only will help determine the stability of network measures as the data is perturbed. This, however, does not fix any other data problems that occurred earlier in the data collection or fusion process.

Reliability Theory-Bayesian Weighting: Lessons learned from reliability and validity in data collection can be used to help assess the validity of network ties and actors in producing data sets (e.g., using Bayesian weighting to better estimate the presence or absence of nodes and ties). Following the work of Romney, Weller and Batchelder (1986) in their work on a formal model (Cultural Consensus Model (CCM)) for estimating the culturally correct answers for cultural data (e.g., beliefs) from patterns of responses from informants, a similar approach can help in better estimating the probability of a node or tie or edge is correct (e.g., present/absent). The basic idea is that for the determination of the presence or absence of ties or nodes there are a number of given

sources of data (i.e., the equivalent of informants in the CCM). Each source provides information on the presence or absence of a given node or edge. There is probably little doubt that the sources will have an overall consensus across all the nodes and edges (fit the consensus model in CCM parlance), but there may be disagreement across sources with respect to some nodes and edges. Following the CCM, data sources that more consistently get the presence or absence correct for highly agreed upon nodes and edges will be weighted more in determining the presence or absence of nodes and edges in the cases of higher uncertainty. This Bayesian approach has worked well for estimating the culturally correct answers in cultural data and seems to have great potential for being adapted for estimating node or edge presence or absence in the social network case when multiple sources of data are involved. So, for example, for a given context there may be multiple sources of data from DoD agencies, from nongovernmental sources, from coalition forces, from signal intelligence and other forms of human intelligence. Each of these sources are like cultural informants in the CCM approach and the "correct" answer can be estimated by aggregating and modeling across sources leading to a much improved means for determining confidence in targeting decisions. Although the formal model at present works only with binary data (e.g., present/absent) the model is being extended to continuous data so that interval level measure and above for edge or tie strength can also be potentially estimated (e.g., the amount of dollars flowing between two actors).

Awareness: No matter how many tools we produce to better determine, estimate and deal with errors in social network data it is important not to loose sight of the critically important role of vigilance and hard work in stopping errors at the very beginning of the data collection process. However, even with such hyper awareness there will inevitably be some error and we should have the proper tools at hand to deal with them.

Discussion and Recommendations

Social network analysis and modeling are powerful tools. But extreme caution should be used in interpreting any analysis unless issues of data quality are seriously considered. This is particularly true for the security context where data errors may have grave implications. Efforts need to focus on developing a science of errors in the network context. This will involve theoretical research exploring and modeling the impact of data errors on network model performance. Further, this will require the development of software tools for conducting error simulations and nodal sensitivity analysis that can be readily used by analysts as well as modelers.

The CCM approach should be adapted for use in social networks error management and mitigation also requiring the adaptation of existing software or the development of new software. In achieving these ends, the enterprise to further advance a science of errors can be funded by such DoD programs as the Multiple University Research Initiative (MURI). These can followed up by the development of software tools by such DoD and other Federal programs with national security focus using STTR or SBIR funding. It is only through a concerted effort in the development of a science of errors in social network data and modeling that application of network analysis can provide a dependable tool for solving national security problems.

References

Bernard, H.R. and P.D. Killworth. 1977. "Informant Accuracy in Social Network Data II." Human Communications Research. 4(1): 3-18.

Bernard, H.R., P.D. Killworth and L. Sailer. 1980. "Informant Accuracy in Social Network Data IV: A Comparison of Clique-Level Structure in Behavioral and Cognitive Data." Social Networks. 2: 191-218.

Bernard, H.R., P.D. Killworth and L. Sailer. 1982. "Informant Accuracy in Social Network Data V: An Experimental Attempt to Predict Actual Communication from Recall Data." Social Science Research. 11:30-66.

Borgatti, S, K Carley, and D Krackhardt. 2006 "On the Robustness of Centrality Measures Under Conditions of Missing Data." Social Networks 28: 124-136.

Boster, J.S., J.C. Johnson and S.C. Weller. 1987. "Social Position and Shared Knowledge: Actor's Perceptions of Status, Role and Social Structure." Social Networks 9:375-387.

Freeman, L.C. and A.K. Romney. 1987. "Words, Deeds and Social Structure: A Preliminary Study of the Reliability of Informants." Human Organization. 46(4): 330-334.

Freeman, L.C and C.M Webster. 1994. "Interpersonal Proximity in Social and Cognitive Space. Social Cognition Vol 12, Number 3:223-231.

Freeman, L.C., A.K. Romney and S.C. Freeman. 1987. "Cognitive Structure and Informant Accuracy." American Anthropologist 89: 310-325.

Johnson J C, J. Boster, and L. Palinkas. "Social Roles and the Evolution of Networks in Isolated and Extreme Environments. The Journal of Mathematical Sociology Volume 27/Numbers2-3, (2003): pp. 89-122.

Johnson J C, J. Boster, and L. Palinkas. "Social Roles and the Evolution of Networks in Isolated and Extreme Environments. The Journal of Mathematical Sociology Volume 27/Numbers2-3, (2003): pp. 89-122. .

Jjohnson, J C and M.K. Orbach. "Perceiving the Political Landscape: Ego Biases in Cognitive Political Networks". Social Networks 24 (2002) 291-310.

Johnson, J C, J.S. Boster, and D. Holbert. "Estimating Relational Attributes From Snowball Samples Through Simulation," Social Networks 11:135-158, 1989.

Killworth, P.D. and H.R. Bernard. 1976. "Informant Accuracy in Social Network Data." Human Organization. 35(3): 269-286.

Killworth, P.D. and H.R. Bernard. 1979/80. "Informant Accuracy in Social Network Data III: A Comparison of Triadic Structure in Behavioral and Cognitive Data." Social Networks. 2: 10-46.

Krackhardt, D. 1987. "Cognitive Social Structures." Social Networks. 9:109-34.

Krackhardt, D. 1990. "Assessing the Political Landscape: Structure, Cognition, and Power in Organizations." Administrative Science Quarterly. 35:342-369.

Kumbasar, E., A.K. Romney and W. Batchelder. 1994. "Systematic Biases in Social Perception." American Journal of Sociology. 100(2): 477-505.

Romney, A.K. and K. Faust. 1982. "Predicting the structure of a communications network from recalled data." Social Networks. 4:285-304.

Romney, A.K. and Susan Weller. 1984. "Predicting informant accuracy from patterns of recall among individuals." Social Networks. 6: 59-78.

Romney, A.K., Weller, S. and Batchelder, W.H., 1986. Culture as a Consensus: A Theory of Culture and Informant Accuracy, *American Anthropologist* 88:313-338.

Lucy Resnyansky, Social Modeling and Simulation for National Security Decision-making: Ethical Challenges

Presented at the DTRA/Sandia Workshop on Challenges in Computational Social Science. Santa Fe, New Mexico. October 25-28, 2010.

ABSTRACT

This paper aims to reveal the ethical challenges that emerge from the use of social Modeling and Simulation (M&S) within a National Security (NS) context, and to understand how they can be addressed. In order to achieve this purpose, the processes of intelligence analysis and decision-making are approached as a socio-cultural activity mediated by tools and embedded into an intersection of professional, organizational, and epistemological cultures. The construction of 'M&S within the NS context' as an object of ethics-focused research draws upon the discussion of the implications of using new Information and Communications Technologies (ICTs) within the NS context, as well as a meta-methodological analysis of the formation of social M&S as a discipline. The ethical challenges which emerge due to the use of modelling and simulation differ from those created due to the use of technologies enabling surveillance and data extraction within the NS environment. In the latter case, the primary concern is the physical or moral harm to citizens, and solutions are sought within the legal regulations area. In the case of analytical activity being supported by modeling and simulation tools, the ethical issues may be less obvious since they relate to the quality of analysis and decision-making within an organization. This kind of ethical challenge can be most effectively addressed in the research, development and assessment stages.

INTRODUCTION

The application of mathematical and computational methods to the analysis of social phenomena is believed to be useful within different areas of practice, including the National Security (NS) domain [1, 2, 3, 4, 5]. For example, social Modeling and Simulation (M&S) can help analysts explore the behavior of cultural groups [6] and political processes [7], and train staff to communicate effectively in cross-cultural interaction [8]. The use of M&S for the analysis of social phenomena seems to be a good way to introduce social science into the NS context. However, the implementation of social M&S may have significant consequences for the NS analysis and decision-making practice. Therefore, a critical reflection on the socio-cultural and ethical implications of this process must become an essential part of the development and implementation of social M&S within the NS area [9]. Due to the specific nature of social M&S, it is difficult to recognise emerging ethical challenges and to locate them within the development-technology-practice nexus. M&S tools are often promoted as a solution to the ethical problems that might emerge in *in vivo* experimental studies of social systems [10]. This view is reflected, for example, in the following definition of Computational Social Science (CSS):

> CSS is an emerging, hybrid discipline that is focused on rendering social theory into computational constructs for the following purposes: To investigate and experiment in situations where direct observation of human behaviour is not possible or not ethical. [11].

Hypothetically speaking, social M&S is not immune to the moral problems that are intrinsic to research which involves human beings. In Stanislaw Lem's *Cyberiada*, two genius robotic engineers are employed to create a model of a perfect world. The constructors simulate a series of micro-societies subject to

computational experimentation, in order to discover a formula of total happiness. They are experimenting with a myriad of combinations of social conditions and variables. They create artificial societies based on all possible sets of values – religious, secular, hedonistic, and so on. The agents' motivations vary from profoundly egoistic to purely altruistic. The most sophisticated mathematical logic and computational algorithms are used. Nevertheless, the simulated social systems ended in all kinds of social catastrophes and tragedies. The constructors conclude that the problem of total happiness has no solution. Finally, the constructors become preoccupied entirely with an ethical problem emerging from the fact that, despite being artificially created and programmed, the agents did experience real suffering.

Those aware of the current developments within the area of social M&S, know that it is far below the level achieved by Lem's characters. It is true that the developers of social M&S tools do not need to worry about moral issues that might arise when humans become an object of research [12]. Also, the user of the M&S tools seems to have no need to worry about ethics. Unlike surveillance, data mining and identification technologies, M&S tools do not intrude into the private lives of citizens and the activities of organizations. These tools are meant for 'internal consumption', they are offered as capabilities supporting the processes of analysis and decision-making. Nevertheless, while being able to provide a solution to some ethical challenges related to the exploration of social and cultural phenomena, social M&S may create new ethical challenges.

The purpose of this paper is to understand what kind of ethical issues are emerging when social M&S is used within the NS area. In order to achieve this, the processes of intelligence analysis and decision-making are approached as a social activity mediated by tools and embedded into an intersection of different professional, organizational, and epistemological cultures.

This study is grounded within a system-activity theory, a logico-philosophical model of interdisciplinary research [13, 14], the concepts of social modeling and Information and Communications Technologies (ICTs)-mediated collaboration as systems of social activity [15. 16], and the concept of technology-practice as an activity involving cultural (values and ethical codes), organizational and technical (knowledge, tools, and resources) aspects [17]. The construction of social M&S as an object of ethics is grounded within the broader research on computing and ethics [18, 19, 20, 21, 22, 23], the implications of new ICTs within the NS context, intelligence analysis and policy making [[24, 25, 26, 27, 28, 29, 30], and a meta-analysis of the emerging disciplinary field of social M&S [5, 9, 31, 32].

ETHICS, TECHNOLOGY, AND NATIONAL SECURITY

Understanding of ethical issues emergent due to the use of technology may vary, depending on the ethical perspective. Ethical issues can be approached either within a teleological perspective (the ethics of ends) or deontological (the ethics of duty) [18]. The teleological perspective on ethical reasoning is represented, for example, by utilitarianism. The deontological perspective is represented by pluralism (a duty-based approach) and contractarianism (a rights-based approach). Ethical reasoning is not entirely subjective. By its nature, it is reasoning in relation to objective criteria, such as basic human rights, maximizing social good, and so forth [18]. At the same time, ethical issues are context- and situation-specific, and, in the case of technology, are determined by the nature of technological tools (Table 1).

Table 3 Information Technology: Ethical issues

Technologies	**Application**	**Ethical issues**	**Solutions**
Technologies enabling access to information (e.g., Internet)	Scholarly publishing Media Entertainment Research Collaboration	Information safety Intellectual property Social injustice: unequal access to information Inappropriate use of knowledge	Legal regulations Economic measures Organizational policies
Technologies enabling surveillance, identification, tracking, and data mining and fusion	National security Intelligence Business	Privacy Citizen rights Democratic values and freedoms Identity fraud Workplace relationships	Legal regulations Law enforcement National policies Organizational policies
Technologies enabling social interaction	Social networks Virtual reality	Privacy False identity Moral values Psychological violence	Legal regulations Codes of ethics regulating intragroup and interpersonal relationships Cultural change
Technologies enabling data analysis and decision-making	Finances Medicine Social work Military operations Intelligence National security Administration Social research	Responsibility Accountability Workplace: loss of jobs Client-vendor relationships	Professional (engineering) ethics Organizational policies

In the case of information technology, the usual foci of ethical examination are vendor-client relations [18]; privacy, civil liberties and rights [20]; social relationships in virtual space [12, 33]; intellectual property and information security [21]; and the use of knowledge in electronic environment [34, 35, 36, 37]. Also, ethical issues emerge due to the impact of new technologies in the workplace, such as: the potential loss of jobs, monitoring of employees' activities, and the reconfiguration of power relations between employees and managers [19, 38]. In the case of expert systems used to support decision-making practices, the accountability and responsibility issue is the major ethical challenge. These practices vary from health care and child-protection risk assessment and emergency services to military operations analysis [38, 39, 40]. One of the consequences of using computational systems in decision-making is that the autonomy of the decision-maker is significantly affected. It is not always possible to understand who should be responsible for the decisions. Is it the user or the manager who influenced the product choice and implementation? Or should the responsibility be located within the design area, because "[c]omputer optimization algorithms like those needed to solve problems in large spaces with many variables and constraints

can only take into account those quantifiable variables that were deemed to be critical in early design stages" [40].

Ethics is linked to such fundamental concepts of market-based economy as quality, trust and contract. Ethics is of enormous importance for IT industry, and is part of the IT curriculum. The main purpose of computer ethics courses is to instruct students in how to address potential ethical and legal issues that may emerge due to the use of information technologies. The problem is that there can be no definite instruction on how to conduct ethical reasoning. Rather, general codes of professional (engineering) ethics are offered, together with specific cases as examples to guide the ethical reasoning. The most practical advice seems to be the following: interpretation of ethical issues depends on how a piece of technology is conceptualized. For example, in order to conduct ethical reasoning regarding the producer's responsibility, it is essential if a piece of software is conceptualized as a *product* or a *service*. In the case of electronic information, court's decisions may differ depending on whether a database was approached as a *book* or a *technology* [18].

The literature on computing ethics is addressed mainly to the IT vendors. Within the business perspective, the main concerns are with liability and fair compensation policies. Solutions for ethical problems are sought within the law/policy/code areas. It is a common perception that the traditional ethical framework can assist in solving the ethical challenges caused by the ICTs.[57] The business-oriented literature on computing ethics highlights the role of managers who make decisions about the implementation of particular capabilities. However, the reasoning behind their decisions seems to be influenced by utilitarian and egoistic motives. The literature highlights corporate loss or manager's career as the consequences of offering low-quality software products and services, and warns against over advertising and overselling products as counter-productive for the industry [18]. Such activities as the design of software, or the choice of 'expert knowledge' to be embodied into expert systems, are practically beyond the realm of literature addressed to the IT industry. In order to position ethics outside the design area, an image is cultivated of software as an extremely complex technology in which bugs and fallacies are unavoidable in spite of genuine efforts to fix them.

Within the national security context, the issue of ethics has been discussed in relation to the use of new technologies enabling identification and surveillance. The implementation of such technologies has generated ethical concerns within different social and professional groups, including civil liberties advocates, regulatory agencies, software developers, health care professionals, the e-commerce community, political scientists, and others [24]. One of the most debated ethical issues is the impact of new technologies on citizen rights and freedoms. This discussion draws upon a legal and political discourse on the nature of the democratic state and its relationships with the individual [25]. It has been argued that new threats make it necessary to re-examine the meaning and value of individual privacy in exchange for promises of safety and security. Also, the concept of privacy has to be re-examined, in order to reflect the realities of the digital society. Accordingly, new legal regulations, organizational policies, and codes of ethics may need to be developed.

Issues such as responsibility and over advertising are also quite relevant within the NS context. ICTs have become a participant in data gathering – hence, in analytical and decision-making processes. Many activities previously delegated to people are now performed by computational systems. For example, the level of danger presented by an individual has previously been assessed by people. In

[57] However, in the case of the Internet, for example, it is technology that is claimed to be a better solution to ethical problems (privacy, in particular) than government regulation [33].

the past, this assessment could be based on the data provided by someone who monitored the 'moral health' of the community residing in geographical proximity; the threat assessment was made by intelligence professionals who identified potentially threatening actors and used up-to-date technology to obtain evidence substantiating their conclusions [41]. The new ICTs are introducing a radically different strategy [42]. It has been argued that the user is encouraged to identify a threatening actor on the basis of evidence extracted from data that can be accessed and processed with the help of the data gathering technological tools. It is, of course, recognized that the power of information technology has to be applied thoughtfully and that technology helps reveal *possibilities* that should be examined and analyzed by humans. However, it is also necessary to remember the logic of technological determinism [33]. ICTs are transforming information consumption on such a fundamental level that their effects can still be hidden from current generations [43, 44, 45, 46, 47, 48]. Technology facilitates access to information about individuals and organizations. However, it may not be the high-quality, meaningful, 'distilled' knowledge about the community that could have been obtained by a human observer [42]. Uncritical reliance on technological capabilities can have a negative effect on NS practitioners' ability to perform their duties and provide a secure environment for the general community [26].

ICTs have provided an unprecedented opportunity for data gathering, which can be of great help in countering threats related to political violence. At the same time, uncritical implementation of new data gathering technologies may result in the range of actors subject to monitoring being restricted to those who have access to a computer and possess a digital identity [49]. The problem of data fusion has not been solved completely – meanwhile, in popular discourses on technology, such as found in the media, the ICTs-enhanced data gathering is often portrayed as a way of knowing everything about everybody [50]. The naturalization of this opinion within the NS context may result in the illusion of security. Therefore, the software engineering community is encouraged to behave responsibly and not offer ineffective measures and tools in the name of counter-terrorism [49]. This is a noble purpose. The question remains: how do we know that an offered solution is going to be effective rather than counter-productive? This question relates to ethics but answering these kinds of questions cannot be a matter of individual moral choice. They should be answered on the basis of a technically sound assessment guided by professional codes of ethics and clearly identified criteria [51]. The codes and criteria, however, need to be linked to specific contexts and technologies. Rather than being approached in terms of a contest between abstract moral and ethical principles and norms, the use of digital technologies within a NS context should be approached as a historically evolving system of social activity characterized by the dynamics of participants' roles and power relations, and shaped by professional ethos, organizational culture, and political conjuncture.

To summarize, the new ICTs can generate different kinds of ethical challenges. Making those ethical challenges visible is a complex task. All the more complex is an understanding of how those challenges can be resolved. Things that may appear as having nothing to do with ethics in the development stage may contribute to the emergence of ethical challenges in the stages of implementation and use. Ethical issues clearly seen in the process of use may appear not to have roots within the development stage. This makes it difficult to develop policies and recommendations catering to specific actors. It is hard to even identify all of the various actors and factors that might contribute to the emergence of ethical problems, or to offer a plausible solution. In order to understand how to address the ethical challenges emerging due to the use of social M&S within the NS decision-making context, the following research tasks need to be completed. First, it is necessary to understand the role of M&S within this context – should it be approached as a product or a service? As a technology or an activity? Second, it is necessary to reveal the range of participants

whose activity and relationships will be affected by the introduction of the M&S tools. Finally, it is necessary to approach the development of social M&S capabilities as a locus of ethics-related issues.

SOCIAL MODELING AND SIMULATION IN A NATIONAL SECURITY CONTEXT – AN INTERSECTION OF CULTURES

M&S has become a recognized professional discipline, which has made it necessary to address the issues emerging from the extended use of M&S within different practices, and to develop a professional code of conduct that would regulate the relationships between simulationists and users [52]. Within the discipline-formation perspective, M&S is approached in terms of "the tasks that are typically carried out by a modeling and simulation practitioner" and in terms of "the lifecycle of a typical modeling and simulation project," and *simulation* is defined as "goal-directed experimentation with dynamic models" [52]. This understanding of M&S reflects the internal specificity of M&S discipline vis-à-vis Physical Sciences, Mathematics, Psychology, and so forth.

This discipline-focused concept of M&S is used as a basis for the formulation of a code of ethics for M&S professionals (i.e., software developers) [53]. The code aims at helping the M&S professionals behave in an ethically responsible way. The principles of professional ethics are supposed to be universally applicable. However, it is difficult to apply those general recommendations and regulations even within those areas in which models and simulations have been used for a long time, such as finance and medicine [54]. It is not clear how to apply them to such a specific case as the development of social M&S tools to support NS analysis and decision-making. In order to be applied within this area, the professional code of ethics needs to be reformulated in terms which reflect the specificity of NS decision-making. Also, in order to answer the ethical challenges emerging due to the use of social M&S within NS area, a piece of technology needs to be approached as an active participant of a system of social activity [16, 55, 56, 57].

The NS area is an intersection of such diverse cultures as social research, engineering, computational science, intelligence business, political decision-making, bureaucracy and management. Depending on the configuration of interacting cultures, different aspects of ethics may become relevant (Figure 1). Ethical awareness can vary from high to low, according to the participants' visions of their roles, their interpretations of NS goals, and their professional ethos and epistemological mindsets. The transformative impact of social M&S results from the fact that it introduces quantitative methodology and computational techniques of information processing and knowledge representation.

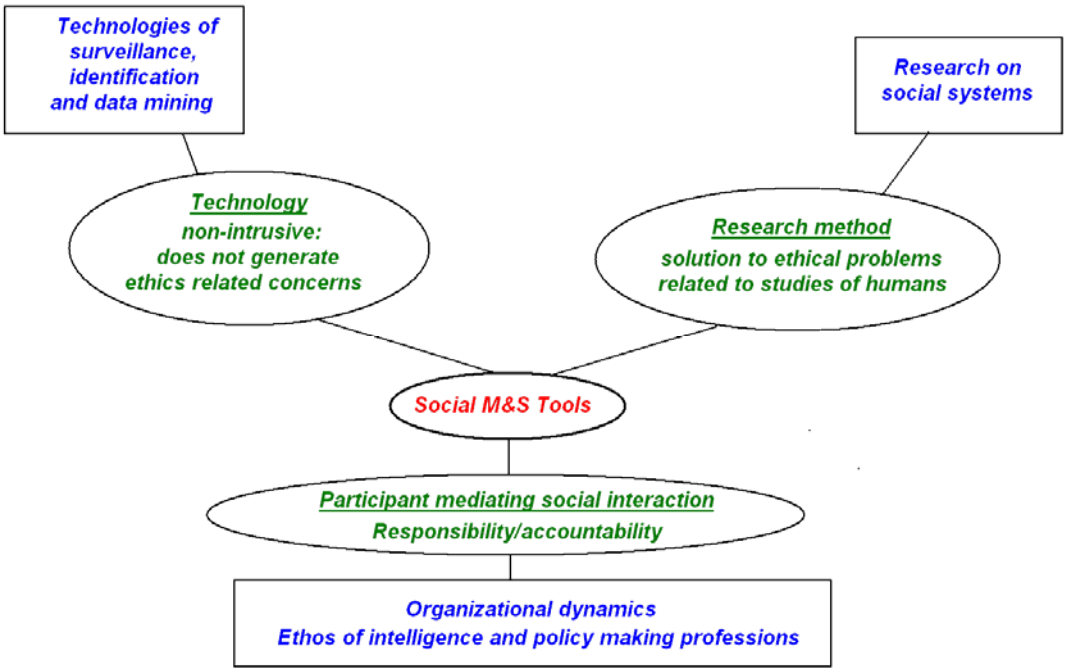


Figure 1 Social M&S within the NS context: a piece of technology, a research method, and a participant of social interaction

In terms of work process, the NS-related analysis and decision-making can be understood as the flow of information. The process can be described by the concept of intelligence cycle [27].

> The intelligence cycle starts with decision-maker information requirements levied on intelligence collection capabilities, the processing of the collected raw intelligence and transmission of this processed material to analysts who decipher its meaning, and relay that understanding back to the decision-makers, who then levy additional requirements. [58, p. 131].

According to intelligence researchers, the primary role for technology within the intelligence area seems to be that of assisting data collection, such as machine translation. Information overload is a good reason for using data mining and aggregation tools. There is, however, a concern that even these kinds of tools may be inadequate and limiting to the analysis [28]. When it comes to analysis, technological tools seem out of place, on the grounds that intelligence analysis is more like an art. It is "an intellectual process based on the application of human thought and judgment" [59, p. 67]. There is a fear that intelligence analysts can be replaced by technology, and that the use of analytic tools may have devastating consequences for the quality of intelligence analysis and, consequently, can negatively affect decision-making.

From the organizational perspective, however, the human nature of intelligence analysis and decision-making may generate serious problems. Currently, the analyst's head is the locus of expertise and institutional memory. At the organization level, this generates such problems as the compartmentalization of knowledge and the failure to share information [60]. Continuous attempts of

organizational reforms proved to be ineffective, due to the paradoxes intrinsic to any organization, such as trade-offs between managerial control and the flow of information:

> Efforts to improve managerial control and lines of accountability [60, p. 126] require boundaries but boundaries block the flow of information. Flexible systems maximize the organization's ability to obtain information but create an environment in which coordination is difficult limiting the effective flow of information. A similar trade-off exists with regard to delegation of authority. Secrecy is highly valued and argues for compartmentalization which, in turn, limits flexibility and information flow while increasing moral hazard and transaction costs. [60, pp. 126-127]

In this situation, technological solutions may appear quite appropriate and attractive. Proposed technological solutions involve, for example, the creation of databases that can become a new locus of expertise and institutional memory. However, the possibility of introducing such technological solutions may be interpreted by analysts as loss of the monopoly of expertise, and may generate uncertainty about tools' impact on the ways in which analysts usually work [59].

The further from the stage of data collection, the less significant seems to be the place assigned to technological tools. The final product of intelligence analysis is the assessment text provided to policy makers to read and base their decisions upon. While intelligence researchers tend to locate analysis in the same semantic field as art and imagination (the activities that can be somehow 'assisted' by technology), the activity of decision-making is clearly separate from this field as being a highly politicized area. Nevertheless, an examination of the intelligence cycle can show that there is a place for technological solutions enabling the process of analysis and, most importantly, enabling intelligence analysts to make the results of their activity more consumable for the decision- and policy-maker.

For example, the analytic process should incorporate alternative ('red team') analyses. Meanwhile, analysts cannot be expected to possess an equal ability to overcome cognitive biases. They may even become captives of their own analytic creations – images of societies or adversaries that they created, which may result in discarding evidences that do not correspond to those images [59]. Texts of assessments should include views that challenge the dominant assumptions. Incorporation of alternative views means that a text of assessment should become, using the literary theory terms, a 'dialogical,' 'polyphonic' text characterized by heteroglossia (simultaneous presence and interaction of multiple voices) [61, 62]. Creating such a text requires outstanding writing skills; the history of world literature knows only a few authors who could work in this genre. Alternatively, the conventions of academic argumentative writing can be used, enabling the author to explicitly identify the boundaries of different approaches exposed within the text [63]. Consuming such kinds of texts requires for the reader to be socialized into advanced literacy cultures. This may not be the case when texts of assessments are consumed within the policy-making area.[58] It is vital to represent analytical results in a way that will make them consumable for decision makers. The analytical results should be reinterpreted in 'layman' terms, presented in a succinct and 'straight to the point' format, and delivered to the addressee in oral or visual mode [26, 64].

> Government intelligence and information systems spend a lot of time paring back large volumes of empirical data into short, digestible forms, gleaning the strategic message for the most senior government officials, and gauging the absorptive capacity of information users at different levels

[58] See, for example, Treverton's chapter 'The intelligence of policy' [26, pp. 177-215], and Gill and Phythian's chapter 'What do they do with the information gathered?' [27, pp. 82-102].

> of government. Aligning the mutual expectations and interactions between information experts and decision-makers remains an elusive goal. [65, p. 39].

Observations on the cultural differences between the intelligence and policy-making communities are of particular relevance for an understanding of the place of social M&S within the NS decision-making area. The relationships within the intelligence-policy nexus, according to some prominent authors in this area, can be better described as 'misconnection'. An impressive picture of the dramatic relationships between intelligence analysts and policy-makers can be found in Treverton [26], who refers to analysts and policy officials as *members of different tribes*. He maintains that intelligence analysts want to think and to understand, not to act. Policy officials are absorbed in here; they come from business or academia to act, to make something happen. They are in the decision-making position for a relatively short period of time, which encourages them to take a shorter view of a problem. Intelligence "is still a written culture, while politics, especially at the top, is mostly oral" [p. 191]. "In these circumstances, most of so-called finished intelligence – that is, analyses established between covers with elegant graphics… – would stay in the pile" [p. 191].

In a succinct overview of the most recent literature on intelligence and NS decision-making, Mark Phythian [59] highlights the following differences. The intelligence community is characterized by resistance to out-sourcing analysis and to bureaucracies. It tries to avoid the risk of politicization because it can compromise the objectivity of their assessments. Policy-makers, on the other hand, tend to form policy preferences on the basis of ideology rather than intelligence analysis, and may support their decisions by evidence from all sorts of sources. Their prioritization of evidence is selective, based on their former experiences and often aims to support their pre-existing views. Therefore, in order to deliver the analytical results to the policy-makers, it is recommended that analysts know more about their customers' background and expertise. The problem is that intelligence analysts do not like "to be entrepreneurs in finding ways to get policymakers to pay attention to their analyses" [26, p. 180].

There is a definite place for such capability as social M&S within this picture. It can become a capability enabling both the exploration of social reality, and the communication of the knowledge to the decision-maker.

Within the intelligence analysis and decision-making processes, the culture of exploration through 'experimentation' needs to be promoted, as numerous examinations of specific cases of intelligence failures have indicated [27, 28, 66, 67]. The affirmation of the value of experimentation within the intelligence and decision-making can provide the modeling community with more definite criteria for the assessment of the social M&S tools as a means enabling explanation rather than prediction [68]. At the same time, it is necessary to become aware of the potential impact of this kind of analytical capability on the NS decision-making. In decision-making, it is important to base one's choice on objective data. Within a NS context, the provision of objective and impartial information is considered the main purpose of intelligence. A rigorous and elaborate system ensuring the reliability and credibility of intelligence information had been developed. At present, however, the intelligence business is undergoing a paradigmatic change [26, 27, 28, 29, 30, 69]. Intelligence is not only about surveillance and data collection from secretive sources. Open sources of data, including the Internet, play an important role in intelligence data gathering and analysis. In the new information environment, the role of intelligence and the relationships within the intelligence-policy nexus have changed. The strict boundaries between intelligence and political decision-making are disappearing. The intelligence profession needs to re-examine the concept of data as objective facts that exist independently from the observer [70].

The nature of new threats and the changing information landscape have made it worthwhile to introduce social science into the NS analysis and decision-making area. At the same time, there is a certain prejudice against 'soft science' methods. This can be explained by the strong influence of the positivist mindset within the intelligence area. Also, this can be explained by the fact that physical sciences are often (and, at times, mistakenly) perceived as an ideal pattern of objective and rigorous analysis. In this situation, social M&S may be perceived as an effective solution. M&S is believed to represent scientific methods in a way that is more recognizable and meaningful to a non-specialist [31]. This makes M&S a good way to introduce scientific rigor into a practice. In particular, M&S may seem a good way to introduce social science into practice by re-representing it with the means of a formalized discourse. This discourse has been portrayed as objective and rigorous within the modern culture [71]. One of the ethics-relevant implications is that the very use of M&S may strengthen the practitioners' belief that their decisions are grounded within a rigorous scientific analysis of objective facts. However, it would be unethical to encourage the idea that sound understanding of socio-cultural processes can be achieved only through the use of the M&S tools. Rather, it is necessary to educate practitioners about the transformative potential of these tools. The practitioners should realize that a profound understanding of social science methodology is needed in order to use social M&S tools.

DEVELOPMENT OF MODELING AND SIMULATION TOOLS: ETHICAL ASPECTS

Unlike more established areas such as M&S for medicine, financial institutions, and management, social M&S for NS is still in its formation stage [9, 15]. This stage is characterized by a negotiation of the roles of natural, computational and social sciences. The foci of discussion include the difference between the object of research in natural and social sciences; the applicability of formal methods to the analysis of social phenomena; and the epistemological status of experimental study vs. cultural and theoretical insights. In spite of the diversity of opinions, the modeling community stands on the same ground regarding the issue that needs to be discussed. It is mainly about 'how' to analyze social phenomena. Here, epistemology and methodology seem to be more relevant than ethics. However, the development of M&S tools for socio-cultural analysis is also a social practice characterized by a re-negotiation of roles and power relations within interdisciplinary teams, and between researchers and developers and different levels of management [9, 72]. Organizational dynamics and interpersonal relationships between members of a multidisciplinary team are manifestations of the competition or collaborative interaction between different fields of knowledge and epistemological traditions. Social M&S requires a profound transformation of knowledge provided by the involved participants. In the process of this transformation, ethical issues related to use of knowledge begin playing a significant part.

The practitioners are interested in 'what' and 'what for' questions, while the developers of M&S tools tend to focus on the 'how' (methods and techniques):

> Just like Galileo exploited the telescope as the enabling instrument for observing and gaining a far deeper and empirically truthful understanding of the physical universe, social scientists and policy analysts should exploit the advanced and increasingly powerful instruments of computation to see beyond the visible spectrum available through the traditional disciplines. [32, p. 260]

This is not to say that the developers ignore the practitioner needs. On the contrary, the common opinion is that the development of models should be governed by the user's needs and purposes.[59] However, it is not easy to implement this general declaration in the form of case-specific research questions and conceptual models. As a result, a model may help answer questions that can be answered within a chosen method rather than those that the practitioner has to answer. On the other hand, blindly following the client's wishes is not the best strategy either. M&S tools need to be client-oriented. However, this does not mean that the client's vision of the problem should remain exempt from critical examination. The so-called 'practitioner needs' are often formulated in terms of general outcomes corresponding to the purpose of project management, rather than in operational terms that could govern the development process.

Within the area of social M&S, it has been gradually acknowledged that the modeling of social phenomena requires an input of social science and subject matter expertise. However, it is easier to acknowledge this than to incorporate social science knowledge into a formal model. Ideally, a good model is believed to be one that unites sound social theory and quantitative (mathematical or computational) technique. In reality, it is usually the method that is the cornerstone of a model. In some papers published under the computational social science banner, the input from social sciences is practically absent – and this is, perhaps, a better case from the perspective of professional ethics (although not from the perspective of the quality of a model as a research instrument). In the worst case scenario, an illusion of such input can be created through a scant incorporation of social theory in an introductory section of a paper, or through including a social scientist into the list of authors. Sometimes, an analysis of computational models of social phenomena can reveal a complete incompatibility between the formal method and the social theory used to construct the object of analysis. For example, a modeling algorithm may draw upon methodology developed for the analysis of the system's change of states. At the same time, the concept of the modeled phenomenon may be grounded within a structuralist paradigm in social science. In order to comply with the professional code of ethics, social modelers need to endorse the use of social science as a genuine source of a formal model. It is an act of professional irresponsibility to simply refer to social science in order to make a model look legitimate.

Ideally, there should be a choice of modeling methods and techniques, and this choice should be governed by the nature of the object. In reality, however, this rarely happens, due to a number of different reasons, amongst them the influence of organizational inertia, funding policies, and individual interests. As a result, only those phenomena that can be modeled within the range of the available techniques are claimed to deserve exploration, which may result in an incomplete and biased analysis negatively affecting NS decision-making. The developers' vision of the social M&S tools is shaped by mathematical and computational approaches. Problems related to intelligence analysis and NS decision-making are reformulated as general classes of problems within those approaches. There seems to be no place for ethical concerns, due to the developers' feeling detached both from the modeled 'reality' and the decisions to be made on the basis of modeling.

The M&S developers may also perceive themselves free from ethical concerns that relate to the marketing of their products. In this case, ethical issues may emerge if the developers oversell their tools, or if they fail to make the tools' limitations visible to the practitioner. The problem is that the

[59] It is, however, the instrumental aspect of computational social science that is usually highlighted and naturalized within a popular discourse on science, such as media coverage of conferences (see, e.g., [73]).

developers may sincerely believe that their tools *are* a good solution, based on their assessments of the methodological rigor and the models' validity. The detachment from the ethical dimension can be explained by the belief that developers deal only with the *how* rather than the *what* of analysis, thus positioning themselves as 'beyond good and evil'. This positioning is amplified when the assessment focuses on the methods and techniques of modeling without incorporating the broader issue relating to the use of quantitative methods for an understanding of social and cultural phenomena. It is also supported by a trend to approach decision-making as a cognitive process rather than a sociocultural activity affected by specific contexts in which it takes place.

In social research, it is important to be open about one's engagement and partiality. It is a methodological requirement to outline the researcher's background, state the reasons behind one becoming engaged in a particular project, and reflect on one's own piece of research as a kind of social practice that aims to invoke a certain social change. Also, an expert reader can understand what theoretical biases and ideological motives have shaped a particular piece of social research. This can be inferred from the list of references, theoretical chapters, and author details. Similarly, the developers of social M&S tools need to be open about the methodological problems related to the use of qualitative and quantitative research within the decision-making and policy formation area. When quantitative methods are used, it is difficult to capture qualitative changes, to represent the entire context of research practice, and to link different levels of analysis. The conceptual linking of different levels of analysis is a fundamental *theoretical* problem within the social science area. Within the modeling area, this issue is approached as a problem that requires a *technical* solution. One of the unfortunate implications of this perception may be that a modeler finds collaboration with a social scientist unnecessary or undesirable.

Addressing ethical challenges at the development stage requires revealing the politics related to the competition between disciplines. In social research, a methodology is closely linked to the object of research [74]. Computational and natural sciences are entering the grounds that have been traditionally within the realm of social science. Nevertheless, there seems to be almost no doubt about the applicability of the introduced epistemological assumptions, including the positivist idea that the method of research can be abstracted from the object of research. The lack of willingness to comply with the epistemological culture of social research is sometimes stated clearly, as within the 'hard science – soft science' discursive opposition. However, in a tacit, yet more systematic way, the competition between disciplines takes place at the institutional level. The mechanisms of scientific communication developed in contemporary science (journals, conferences, reviewing, and so on) can be used in order to keep qualitative social research outside the magic circle of formal approaches and rigorous methods. Journals may reject papers that aim to critically approach the entire project of social M&S, under the pretext that those papers are not focusing on 'technical content'. This kind of social practice contributes to the naturalization of the idea that the method is separate from the object and the context in which it is going to be used. However, this is not the most productive strategy. At present, social M&S is closer to what Thomas Kuhn [75] calls a 'revolutionary science'. This field "is relatively immature, particularly when compared to the use of computers to construct models of physical and biological phenomena" [9, p. 1]. It is unethical to use the powerful machinery of the institution of science to create an image of social M&S as a 'normal science'.

The development of social M&S tools should not be distanced from areas of ethics and moral responsibility. On the contrary, the ethical aspect becomes more prominent due to the power of technology and the new challenges within the NS area. However, even if developers genuinely care about social and ethical issues, it may not be possible to link methodological concepts with ones that

belong to the realm of philosophy and social and legal sciences (ethics). This makes it particularly important to develop an ethics-oriented assessment and evaluation framework. Evaluation is the area in which it is possible to merge the engineering and the practitioner perspectives, since it requires consideration of the tool in a specific context of use. Specifically, it is necessary to address the ethical challenges emerging due to the difference in the interests of the practitioner and the modeler. Figure 2 shows the place of critical reflection in the provision of high-quality M&S tools.

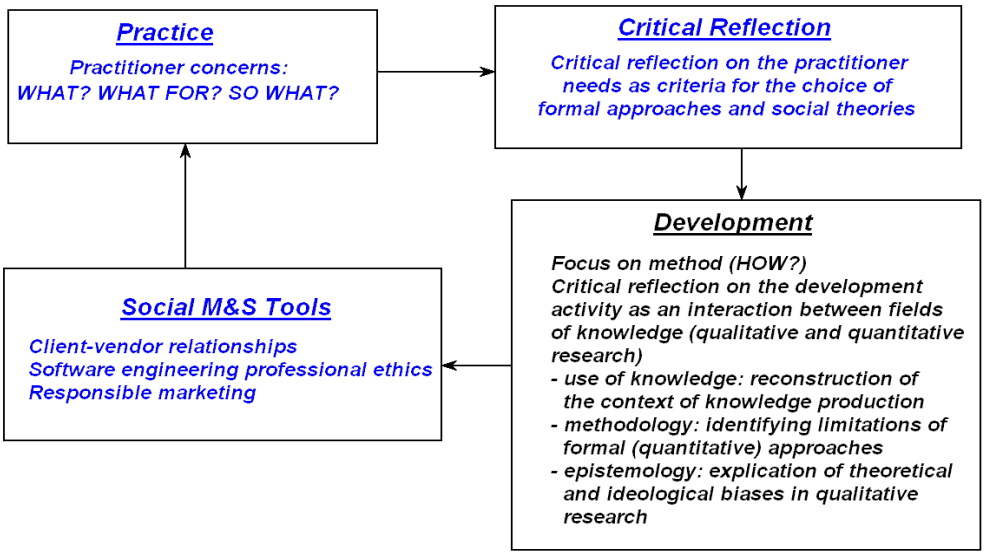


Figure 2 Development of social M&S tools: the role of critical reflection

For the developers, adoption of an ethically responsible position implies that the development of social M&S tools is to be shaped by the needs of a particular user rather than by the questions that a given modeling approach or a specific social discipline allow to ask. The needs and questions relevant to a particular area of practice should be conceptualized as the criteria for the choice of a modeling approach, an assessment of the heuristic significance of the representations of the object created in a particular piece of social research, and the evaluation of the models. However, the developers should not passively follow this practice. Rather, a critical reflexive analysis of the practitioner's needs should be conducted. A critical reflexive stance should also be adopted by the developers of social M&S tools in relation to their own activity [15; 16]. The developers of M&S tools need to:

- identify those formal categories into which qualitative social models are translated;
- make explicit the process of reinterpretation of the qualitative representations of the object into formal categories; and
- provide foundations for their choice of mediating concepts.

CONCLUSION

Social M&S can become an invaluable means for considering the complex problems that intelligence analysts and decision-makers encounter within the contemporary environment. However, there is a danger that the M&S tools will become just a façade that allows analysts and decision-makers to pretend that their decisions are grounded within social science.

In order to address the ethical challenges emerging due to the implementation of social M&S tools in the NS analysis and decision-making practices, it is necessary to identify the relevant differences between the interacting cultures. This task requires that social M&S tools are conceptualized appropriately. Currently, the social M&S tools are defined either in relation to other kinds of technological tools, or in relation to mathematical and computational approaches. This paper suggests that social M&S needs to be conceptualized as a 'participant' that mediates social interaction shaped by different organizational, professional, and epistemological cultures. This approach helps reveal the following key problems:

- Social M&S are able to introduce qualitative social research into the analysis and decision-making processes. This might contribute to the deeper understanding of the object of analysis. However, theoretical and ideological biases intrinsic to social research may become invisible, due to the formalization of social science knowledge during the course of modeling;
- Social M&S allows the use of quantitative methods for an analysis of social phenomena. On the one hand, this enables a more rigorous and comprehensive analysis. On the other hand, there is a possibility of offering incomplete and decontextualized representations of the objects of analysis, due to the limitations of quantitative methodology in studies of social reality;
- Social M&S is an advanced technology of knowledge representation. At the same time, it is capable of re-interpreting analytical findings and is very demanding of the quality of data to be processed.

The development of social M&S tools needs to be grounded within a rigorous scientific approach. However, the use of quantitative methods and computational techniques does not guarantee scientific rigour. It is necessary to make sure that an adequate social science theory is chosen, that it corresponds to the mathematical and computational approaches, that it has been implemented in the conceptual model of a phenomenon, and that this conceptual model has not been distorted in the final product. Also, the development of social M&S tools needs to be reinforced by a clear understanding of the needs and concerns that may emerge in the stages of implementation and use, including those related to ethical aspects and sociocultural implications.

In the transformed, science-saturated practice of NS analysis and decision-making, the cultures of social research, scientific exploration and engineering (which are not easily compatible with each other) are introduced into the intelligence and bureaucratic-political environment. Social M&S tools are becoming a locus of intersection of very different values and sources of authority: expert opinion and scientific truth versus administrative power and political conjuncture, the formal logic of computational algorithms versus the art of intelligence analysis, the universal quantitative methods versus theoretical insights based on case-specific qualitative explorations. There is a chance that the models of thinking offered together with the social M&S tools will be rejected by the culture of an organization, or, that the introduction of the technology will amplify the influence of the bureaucratic and technocratic approaches.

LIST OF REFERENCES

[1] D. Patil, J. Perry, and D. Hamon, *Threat Anticipation Project Overview*, Advanced Systems and Concepts Office, Defence Threat Reduction Agency, 2005.

[2] *Threat Anticipation: Social Science Methods and Models, The 1st Annual Joint Threat Anticipation Center Workshop, 7-9 April*. The University of Chicago, 2005.

[3] L. Resnyansky, *Integration of Social Sciences in Terrorism Modeling: Issues, Problems and Recommendations*, DSTO-TR-1955 (U), Commonwealth of Australia, 2007.

[4] C. Cioffi-Revilla and S.P. O'Brien, "Computational Analysis in US Foreign and Defence Policy," in *ICCCD 2007 – Proceedings of the First International Conference on Computational Cultural Dynamics,* D. Nau and J. Wilkenfeld, Eds. Menlo Park, CA: AAAI Press, 2007, pp. 26-36.

[5] *Behavioral Modeling and Simulation: From Individuals to* Societies, G.L. Zacharias, J. MacMillan, and S.B. Van Hemel, Eds. Washington, DC: The National Academy Press, 2008.

[6] A. Sliva, M.V. Martinez, G.I. Simari, and V.S. Subrahmanian, "SOMA Models of the Behaviors of Stakeholders in the Afghan Drug Economy: A Preliminary Report," *ICCCD 2007 – Proceedings of the First International Conference on Computational Cultural Dynamics*, D. Nau and J. Wilkenfeld, Eds. Menlo Park, CA: AAAI Press, 2007, pp. 78-85.

[7] J. Ozik, D.L. Sallach, and C.M. Macal, "Modeling Dynamic Multiscale Social Processes in Agent-based Models," *IEEE Intelligent Systems*, vol. 23, no. 4, pp. 36-42, 2008.

[8] C.A. Miller, P. Wu, and H.B. Funk, "A Computational Approach to Etiquette: Operationalizing Brown and Levinson's Politeness Model," *IEEE Intelligent Systems*, vol. 23, no. 4, pp. 28-35, 2008.

[9] J.G. Turnley and A. Perls, *What is a Computational Social Model Anyway?: A Discussion of Definitions, a Consideration of Challenges, and an Explication of Process*, Report No. ASCO 2008-013, Advanced Systems and Concepts Office, Defense Threat Reduction Agency, 2008.

[10] J. Johnson, "Science and Policy in Designing Complex Futures," *Futures*, vol. 40, no. 6, pp. 520-536, 2008.

[11] R. Goolsby, "Computational Social Science, Culture and the Global War on Terror," Office of Naval Research Presentation, 2005, www.au.af.mil/au/awc/awcgate/.../onr_soc_sci_culture_gwt_05jul.pdf.

[12] M. Quigley, "Computers and Society," in *Ethics in ICT: An Australian Perspective*, D. McDermid, Ed. Pearson Education Australia, 2008, pp. 249-276.

[13] G.P. Shchedrovitsky, *Methodological Organization of System-Structural Research and Development: Principles and General Framework*, 2005-2007, http://www.fondgp.ru/ lib/int/5.

[14] G.P. Shchedrovitsky, *Methodological Problems of System Research*, 2005-2007, http://www.fondgp.ru/lib/int/5.

[15] L. Resnyansky, "Social Modeling as an Interdisciplinary Research Practice," *IEEE Intelligent Systems*, vol. 23, no. 4, pp. 20-27, 2008.

[16] L. Resnyansky, "The ICTs-mediated Collaboration as a System of Social Activity," *Proceedings of the 2010 IEEE International Symposium on Technology and Society: Social Implications of Emerging Technologies, 7-9 June 2010, University of Wollongong, New South Wales, Australia*, 2010, http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=05514646.

[17] A. Pacey, "Technology: Practice and Culture," in *Controlling Technology: Contemporary Issues*, W.B. Thompson, Ed. Buffalo, NY: Prometheus Books, 1983/1991, pp. 65-75.

[18] R.A. Spinello, *Ethical Aspects of Information Technology*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[19] J.M. Kizza, *Ethical and Social Issues in the Information Age*. London: Springer, 2007.

[20] L.J. Pourciau, Ed., *Ethics and Electronic Information in the Twenty-first Century,* West Lafayette, Indiana: Purdue University Press, 1999.

[21] K.W. Bowyer, *Ethics and Computing: Living Responsibly in a Computerized World*. Piscataway, NJ: The Institute of Electrical and Electronics Engineers, 1996.
[22] M.J. Quinn, *Ethics for the Information Age*. Boston: Pearson Addison Wesley, 2006.
[23] C. Zielinski, P. Duquenoy, and K. Kimppa, Eds., *The Information Society: Emerging Landscapes. IFIP International Conference on Landscapes of ICT and Social Accountability, Turku, Finland, June 27029, 2005*. New York, NY: Springer, 2006
[24] J. Waldo, H.S. Lin, and L.I. Millett, Eds., *Engaging Privacy and Information Technology in a Digital Age*. Washington, D.C.: The National Academy Press, 2007.
[25] L.S. Strickland, D.A. Baldwin, and M. Justen, "Domestic Security Surveillance and Civil Liberties," *Annual Review of Information Science and Technology, 39*, B. Cronin, Ed. Medford, NJ: Information Today, 2005, pp. 433-513.
[26] G.F. Treverton, *Reshaping National Intelligence for an Age of Information*. Cambridge, UK: Cambridge University Press, 2003.
[27] P. Gill and M. Phythian, *Intelligence in an Insecure World*. Cambridge, UK: Polity Press, 2006.
[28] G.F. Treverton and C. Bryan Gabbard, *Assessing the Tradecraft of Intelligence Analysis*, RAND Corporation, 2008.
[29] P. Berkowitz, Ed., *The Future of American Intelligence*. Hoover Institution Press, 2005.
[30] L. Resnyansky, "The Internet and the Changing Nature of Intelligence," *IEEE Technology and Society Magazine*, vol. 28, no. 1, pp. 41-47, 2009.
[31] J.M. Epstein, *Generative Social Science: Studies in Agent-based Computational Modeling*. Princeton and Oxford: Princeton University Press, 2006.
[32] C. Cioffi-Revilla, 'Computational Social Science', *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 3, pp. 259-271, May/June 2010.
[33] R.A. Spinello, *Cyberethics: morality and law in cyberspace*. Sudnury, MA: Jones and Bartlett Publishers, 2006
[34] L.A. Lievrouw and S.E. Farb, "Information and Equity," *Annual Review of Information Science and Technology: Vol. 37*, B. Cronin, Ed. Medford, NJ: Information Today, 2003, pp. 499-539.
[35] B. Nelson, "Empty Archives", *Nature*, vol. 461, pp. 160-163, 2009.
[36] C. Beardon, 'Information Cultures and Counter-cultures', *The Information Society: Emerging Landscapes. IFIP International Conference on Landscapes of ICT and Social Accountability, Turku, Finland, June 27-29, 2005*, C. Zielinski, P. Duquenoy, and K. Kimppa, Eds. New York, NY: Springer, 2006, pp. 3-22.
[37] J.J. Britz and S.R. Ponelis, "When Is It Good to Steal? A Moral Reflection on Current Trends in Intellectual Property," *The Information Society: Emerging Landscapes. IFIP International Conference on Landscapes of ICT and Social Accountability, Turku, Finland, June 27-29, 2005*, C. Zielinski, P. Duquenoy, and K. Kimppa, Eds. New York, NY: Springer, 2006, pp. 23-36.
[38] R. Vuokko, "Work Ethics and Effectiveness in Home Care: Rationalization through Technological Innovation," *The Information Society: Emerging Landscapes. IFIP International Conference on Landscapes of ICT and Social Accountability, Turku, Finland, June 27-29, 2005*, C. Zielinski, P. Duquenoy, and K. Kimppa, Eds. New York, NY: Springer, 2006, pp. 281-302.
[39] D. Wallace, "Practitioner Accountability and Decision-making Technology," *Proceedings of the 2010 IEEE International Symposium on Technology and Society: The Social Implications of Emerging Technologies', 7-9 June 2010. University of Wollongong, New South Wales, Australia*, 2010.
[40] M.L. Cummings and S. Bruni, "Collaborative Human-Computer Decision-making in Network Centric Warfare," Paper presented at the *TTCP HUM TP7. Workshop on Aerospace Human Factors Issues in Network-Centric Warfare*, Salisbury, UK, 2005; available at web.mit.edu/aeroastro/labs/halab/papers/IJAPCummings_TTCP.pdf.
[41] W. Diffie and S. Landau, "Brave New World of Wiretapping," *Scientific American*, vol. 299, no. 3, pp. 56-63, 2008.

[42] K.M. O'Connell, "The Role of Science and Technology in Transforming American Intelligence," *The Future of American Intelligence*, P. Berkowitz, Ed. Stanford, CA: Hoover Institution Press, 2005, pp. 139-174.
[43] N.C. Burbules and T.A. Callister, Jr., *Watch IT: The Risks and Promises of New Information Technologies for Education*. Boulder, CO: Westview Press, 2000.
[44] T. Roszak, *The Cult of Information: The Folklore of Computers and the True Art of Thinking*. New York: Pantheon Books, 1986.
[45] D. Rushkoff, *Children of Chaos: Surviving the End of the World as We Know It*. London: Harper Collins, 1997.
[46] G. Riva and C. Galimberti, Eds., *Towards Cyberpsychology: Mind, Cognition and Society in the Internet Age*. Amsterdam: IOS Press, 2001.
[47] J.P. Gee, *Power Up: What Video Games Have to Teach Us about Learning and Literacy*. London: Palgrave/Macmillan, 2003.
[48] M. Poster, *The Information Subject*. Amsterdam: G+B Arts International, 2001.
[49] M.A. Caloyannides, "Engineering or Sloganeering? The Counterattack on Privacy," *IEEE Security and Privacy*, March/April, pp. 84-87, 2003.
[50] S. Garfinkel, "Information of the World, Unite," *Scientific American*, vol. 299/3, pp. 82-87, 2008.
[51] S.H. Unger, *Controlling Technology: Ethics and the Responsible Engineer*, New York: John Wiley, 1994.
[52] L.G. Birta, "The Quest for the Modeling and Simulation Body of Knowledge," the *Sixth Conference on Computer simulation and Industry Applications*, Instituto Technologico de Tijuana, February 19-21, 2003, www.sim-summit.org/simsummit/0303/0303fa.pdf.
[53] T.I. Ören, M.S. Elzas, I. Smit, and L.G. Birta, "A Code of Professional Ethics for Simulationists," *Proceedings of the 2002 Summer Computer Simulation Conference*, San Diego, CA, July 13-18, 2002, www.site.uottawa.ca/~oren/pubs/D81_Code.pdf.
[54] A.L. Tongen and M.B. Adam, "Ethics Involved in Simulation-based Medical Planning," *Ethics and Medicine*, vol. 22, no. 1, pp. 23-30, 2006.
[55] P. Bennett and L. Resnyansky, "How the Concept of Ethnicity Can Inform Our Understanding of the Potential Impact of Security-related Technology upon Work Practices and Society," in *Recent Advances in Security Technology: Proceedings of the 2006 RNSA Security Technology Conference 19-21 September, Canberra*, P. Mendis, J. Lai and E. Dawson, Eds., pp. 143-158, 2006.
[56] B.A. Nardi, Ed., *Context and Consciousness: Activity Theory and Human-computer Interaction*. Cambridge, MA: The MIT Press, 1997.

[57] H. Hasan, I. Verenikina and E. Gould, Eds., *Information Systems and Activity Theory: Volume 3. Expanding the Horizon*. The University of Wollongong Press, 2003.
[58] S. Marrin, "Intelligence Analysis and Decision-making: Methodological Challenges," *Intelligence Theory: Key Questions and Debates*, Gill, P., Marrin, S. and Phythian, M., Eds. Routledge: London and New York, 2009, pp. 131-150.
[59] M. Phythian, "Intelligence Analysis Today and Tomorrow," *Security Challenges,* vol. 5, no. 1, pp. 67-83, 2009.
[60] G.P. Hastedt and B.D. Skelley, "Intelligence in a Turbulent World: Insights from Organization Theory," *Intelligence Theory: Key Questions and Debates*, P. Gill, S. Marrin and M. Phythian, Eds. London and New York: Routledge, 2009, pp. 112-130.
[61] M.M. Bakhtin, *Problems of Dostoevsky's Poetics*, trans. C. Emerson. Minneapolis: University of Minnesota Press, 1984.
[62] M.M. Bakhtin, *Speech Genres and Other Late Essays*, trans. V. McGee. Austin, TX: University of Texas Press, 1986.

[63] M.A.K. Halliday and J.R. Martin, *Writing Science: Literacy and Discursive Power*. London: The Falmer Press, 1993.
[64] E.R. Tufte, *The Cognitive Style of PowerPoint*. Chesgire, CT: Graphics Press, 2003.
[65] *The Information and Intelligence Needs in the 21st Century Threat Environment*, The Henry L. Stimson Center, Report No. 70, Washington DC, September 2008,
[66] R.K. Betts, "Analysis, War, and Decision: Why Intelligence failures are Inevitable," *Intelligence Theory: Key Questions and Debates*, P. Gill, S. Marrin and M. Phythian, Eds. London and New York: Routledge, 2009, pp. 87-111.
[67] C. Vandepeer, "Intelligence Analysis as Decision-making: A Case Study of the 2002 Bali Bombings," paper presented at *Oxford Intelligence Forum*, January 22, 2009, http://goa-intelligence.org/main/wpcontent/uploads/2009/01/intelligence_analysis_as_decisionmaking_2002_bali_bombing_case_study.pdf.
[68] J.M. Epstein, "Why Model?," *Journal of Artificial Societies and Social Simulation*, vol. 11, no. 412, http://jasss.soc.surrey.ac.uk/11/4/12.html.
[69] P.H.J. Davies, "Intelligence, Information Technology, and Information Warfare," in *Annual Review of Information Science and Technology, 36*, B. Cronin, Ed. Medford, NJ: Information Today, 2002, pp. 313-352.
[70] G.J. Schmitt, "Truth to Power? Rethinking Intelligence Analysis," in *The Future of American Intelligence*, P. Berkowitz, Ed. Hoover Institution Press, 2005, pp. 41-64.
[71] S. Gerovitch, *From Newspeak to Cyberspeak*. Cambridge, MA: The MIT Press, 2002.
[72] E.P. MacKerrow, "Agent-based Computer Simulation for Operational Risk Analysis," *Eclipse Modeling Days*, November 16, Credit Suisse, New York, 2009, MacKerrow091116AgentBasedComputerSim[1].pdf.
[74] S. Weinberger, "Can Social Scientists Win the War on Terrorism?" *Wired,* August 8, 2007, http://www.wired.com/dangerroom/2007/08/in-short-what-a/.
[75] H. Suganami, "Narrative Explanation and International Relations: Back to Basics," *Millenium – Journal of International Studies*, vol. 37, no. 2, pp. 327-356, 2008.
[76] T.S. Kuhn, *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press, 1962.

David Sallach, Social Science Model Validation: Domain-Specific Challenges[*]

Presented at the DTRA/Sandia Workshop on Challenges in Computational Social Science. Santa Fe, New Mexico. October 25-28, 2010.

The social science domain is complex, high-dimensional and dynamic. Many issues concern the intentionality of social actors at multiple scales while they engage in dynamic interaction. Indeed, these social actors may have an interest in avoiding being predictable to other actors. Outcomes are interactive in complex ways and, thus, regularly are at least partially unintended. Under these demands, the validation of social science models definitely constitutes a significant challenge.

The focus of this discussion is the validation of large, complex social models in historical settings. The exact purpose of such models will vary. In some cases stakeholders seek a predictive model. More frequently, the objective is a 'course of action' model that explores the diversity of possible and/or likely consequences. Whatever the specific goal of the model, validation is a *process* through which various criteria (e.g., accuracy, robustness, coherence, etc.) can be assessed.

Validation takes many forms. In some cases, experimental testing is possible. In other settings, conceptual validation is more useful. For social models, specialists argue that complexity ensures that validation will always be incomplete (Hartley & Starr 2010:312). The present analysis examines issues of validation from a social theoretical perspective, identifying a number of sources of validation challenges inherent in the application of agent-based or hybrid models to complex social phenomena. In so doing, it introduces additional criteria for evaluating complex social models.

[*] The author is grateful to Jennifer Perry and Harriet Sallach, who read the manuscript and provided helpful feedback. I also appreciate insights of the participants in the Challenges in Computational Social Science workshop organized by Laura McNamara.

As a preliminary example, even under the most formal of social relations (e.g., enacted legal codes and strong, pervasive norms), rules require definition of the circumstances under which they are applicable, and this issue is commonly and repeatedly contentious. Moreover, rules evolve over time, and in response to a diverse mix of social forces, some of which are endogenous to the social process in question (Heritage 1984:103-134; Burns & Flam 1987).

In sum, the identification and representation of stable social rules does not define an effective strategy for constructing reliable social models, or achieving a baseline for validating them. After reviewing a variety of domain-specific challenges to the validation of social models, we will explore the robustness criterion as a promising basis for social model validation.

Much has been learned in validating and verifying models in the natural sciences, and this knowledge should be applied to the validation[60] of social science models whenever it appears to be appropriate. However, to the extent that these techniques appear not to address the unique issues of the social domain, the incentive to apply them is reduced. They could even provide a source of false confidence in the applicability and usefulness of the model. More likely, they will be recognized as not fully addressing the validation issues inherent in complex social models, and therefore lack the intended effect. Either way, such insights need to be deepened in domain-specific ways.

The present paper will introduce and consider a number of ways in which the social sciences present unique methodological challenges to effective model validation. These include: 1) the high and fluid dimensionality required to adequately represent social models, 2) the way social factors can be mitigated or intensified by the social and/or historical context in which they appear, 3) the situated way in which communications and actions are interpreted by social actors (cultural indexicality), 4) the fact that social factors take form and have influence within the model itself

[60] Although verification and validation are closely linked in the literature, computational social models provide few unique issues relative to natural science models, on the one hand, and the best practices of software development on the other. Accordingly, the present discussion restricts itself to validation issues.

(endogeneity), rather than simply being manipulated by the researcher through model design and configuration, and 5) the inaccessibility of the motives and intent of social actors. This list does not include the already difficult issues arising from the parallel application of complexity theories (Sallach 2000a), which is beyond the scope of the current discussion.

NON-LINEARITIES

It is safe to say that social processes are not reliably, stably linear. Linear models can provide the basis for some social apparent predictability, but their limitations carry a heavy cost. Precisely when changes, including potentially significant structural changes, are imminent, the efficacy of a linear model disappears. This is one reason why models must ultimately be grounded in domain-specific social processes, rather than linear computational representations.

However, to say that social phenomena are susceptible to non-linearities, is not sufficient. As the discussion to follow, demonstrates, there are many social sources of nonlinearity. The focus of the present section is to explore various forms of non-linearities, how social manifestations might be represented, and how models that address these issues might be validated.

As an introduction to social non-linearities and their consequences, consider Abbott's (1988) summary of the fundamental assumptions of social non-linearities. In brief, these include: 1) fixed entities with attributes, 2) monotonic causal flows, 3) univocal meanings, 4) the absence of sequence effects, 5) casewise independence, and 6) independence of context. Each of these assumptions has important implications and, in interaction, they capture many of the sources on social non-linearity. Accordingly, this one article serves as a useful introduction to the validation challenges arising within, and as a result of, the social domain.

Fixed Entities. The assumption that social entities remain the same over a given period of time (or correlative model run) is a convenient one, but it is often misleading. The various nations,

institutions, firms, movements, insurgencies, etc., that researchers seek to model have diverse histories, structures and cultures. To generalize, without taking such differences into account, is to fail to fully engage the domain. As Abbott points out, even considering only creation, destruction, amalgamation and division as the events of interest, each event changes the frame within which inferences (logical or statistical) can be made. Changes in the composition and sampling of social actors may also produce a similar limit on generalizability.

Abbott suggests the subject-event model of history (and narrative) as an alternative framework to a highly generalized analytical model. The central subject (which may be an entity, a structure or a process) is shaped by the events it undergoes. But, the very diversity of possible subjects and potential events suggests that such a framework might accentuate modeling and validation issues.

However, the limitation of the 'fixed entity' assumption may be addressed in other ways. Defining richer models that represent the ways in which social entity types vary, and under what circumstances. Thus, while developing effective domain-specific models that are also theoretically grounded may be challenging, the effort will avoid over-simplified assumptions and will, over time, result in models and applications that are more susceptible to validation.

Causal Flows. The assumption of unchanging causal flows projects causal effects forward in time, and/or downward in scale without fluctuation (or other dynamics). Similarly static assumptions include a single time horizon of social actors, and the continual relevance of particular causes. Abbott provides examples in which causal factors operate on diverse time scales and are activated by contexts and contingencies. As Abbott (1988:174) writes:

> The problem with the whole approach is that the values of these measures at any given time are not freely variable … [I]ndependent variables don't really stand for [a] state's free expression of its intents, but rather for what it can intend *given* the various events it finds itself within. One could imagine measuring these events with moving averages, but the 'width' of the moving averages would have to change with the temporal duration of the events involved … [L]inkages of various yearly levels of variables into larger 'events' undermine studies assuming uniform time-horizons …

Univocality. The assumption of univocal meaning implies that concepts (and variables) have the same meaning and comparable effects throughout the study: in different times, regions, settings and cultural climates. Abbott identifies a number of technical work-arounds but, for present purposes, it is sufficient to note that failure to recognize that an unrecognized multiplicity of meanings may compromise the validity of a study by failing to capture shifts in the interpretive framework used by actors within the model.

Sequence Effects. The assumption that the order of events has no influence on outcomes further adds to the fragility of general linear models. The need is for "methods that can classify or cluster sequential data, such as the histories of individuals, occupations and revolutions" (Abbott 1988:178). Such clustering would provide a means of recognizing, and taking into account, path-dependent effects.[61]

Independence of Cases. Linear models also assume that historical cases are independent of each other. While the importance of relevant data is widely recognized in the modeling community, there are complicated data issues that often remain unaddressed.[62] The immediate issue Abbott raises concerns correlated factors, variables that may together give rise to 'emergent attributes' or syndromes. The inherent complexities of such data clustering (that, in turn, represent substantive relationships) carry validation issues.

A number of additional data issues need to be assessed as well. Archival data is collected for purposes that are not closely tied to the research at hand. Data categories may be distorted or compromised by collection and/or analysis procedures (cf., Garfinkel 1967:188-207). Observable data are often used as a proxy for data that cannot be directly collected, and the assumptions underlying such a mapping must be carefully considered in any modeling application.

---

[61] See pages 12-14 below for a more specific discussion of path-dependent effects.
[62] For a thoughtful treatment of the relationship between historical cases and how they may be drawn upon to shape situated social theories, see George & Bennett (2005).

Variable Independence. A final assumption of linear models that Abbott identifies is that each variable should be independent. "Its effect does not change as other variables change around it, nor is its causal effect redefined by its own past" (Abbott 1988:180). There are technical means of handling various types of factor interdependence, but they have drawbacks of their own. Since policy-oriented and historical processes are defined by rich interactions among social actors (and among the variables that serve as their proxies), such complexities must be addressed in assessing the models used to represent them.

Abbott assumptions well as a preliminary checklist for analysts seeking to confront the unique requirements of social models. However, to fully address social validation issues, greater depth is required. There are facets that require further exploration, and issues not yet addressed. It is to these topics that we now turn.

Contexts and Complexities. Table 1 provides a brief description of each domain characteristic, and the implications they have for the validation of social models. In the discussion to follow, these issues have been categorized as examples of either social complexities or context effects, and are explored in greater depth.

SOCIAL COMPLEXITIES

The complexities of the social domain can be expressed in a variety of ways. First, consider the distinction between factors that interact within a model and those that are external, with their causes lying outside of the analytical system.

Endogeneity. Whether an attribute is exogenous or endogenous is usually regarded as a characteristic of the *model* rather than the social process that is being represented. However, the structure of a model necessarily reflects assumptions

Domain Characteristics	Characteristic Descriptions	Validation Implications
Endogeneity	Social actors are changed by intra-model processes and by social ecologies.	Structures and propensities of social actors can't be stipulated once and for all, but are in flux.
Social Dimensionalities	Social phenomena are complex; representing them requires spaces with high dimensionality.	With many interactions and potential outcomes, the possibility space of complex social processes is vast.
Trans-scale Propensities	Social phenomena occur at many scales from the nuclear family to the UN, and from a moment to a millennium.	Social processes at multiple levels interact with each other, shaping contexts, creating many complex and indirect causal flows.
Interaction, Paths and Strategies	Social interaction is an endogenous and self-referential process. It creates fine-grain sequential paths and dependencies.	Small differences in interactive processes result in significantly different paths. Similar initial conditions can produce strikingly divergent results.
Interpretive Agency	Social actors are oriented by meanings, both individual and shared, shaped through interpretive skills.	Since changed interpretations may result in altered courses of action, many social models must model meaning attribution.
Inferring Intent	Given mixed motives and deception, there is no direct evidence of actor intent. As in life, intent must be inferred.	Robust actor models involving both reason and affect may be necessary in order to explore patterns of motive and intent.

Table 1. Domain Characteristics and Validation Implications

about the domain. A critical issue that the 'exogeneity versus endogeneity' choice concerns is the complexity of the causal field. In general, the fewer the endogenous variables, the simpler and less interactive the model is, and the less likely that it will adequately represent complex social processes.

In contrast, capturing even simple forms of endogenous interaction can lead to more subtle and compelling models. As an early economic example, Kurihara (1960) rejected the idea that growth cycles were produced by exogenous shocks, substituting a model that better captures the dual character of capital by generating self-limiting oscillations around the steady growth path of a dynamic equilibrium. Together, demand-decreasing and capacity-increasing effects are counterposed, thereby generating both a positive rate of growth, and limit cycles that accompany and envelope it. It is unlikely that an exogenously driven model could achieve a comparable level of elegance and coherence.

In a more methodological example, Wlezien, Franklin & Twiggs (1997) demonstrate that a failure to identify endogenous cross-effects results in an exaggeration of the extent to which economic perceptions determine voting choices. In this case, the point is that voters tend to take a

favorable view of the economic policies of candidates they prefer. The authors accurately describe the required analytical process as one of disentangling endogenous effects.

A similar point has been made regarding 'political business cycles', in which researchers had suggested that politicians were able to use economic stimulation as a means of improving their chances for reelection. However, when the empirical patterns appeared murky, Heckelman and Berument (1998) noted that there had been an operative assumption that the time of election occurrence was fixed. However, since many parliamentary systems allow the sitting government to have flexibility regarding when an election will be scheduled, the relationships between economic trends and elections were more complex than had been previously thought. Specifically, the effects could flow either way with either political or economic factors having causal priority.

Two further examples illustrate the potentially broad significance that socially endogenous interactions may have. Sunstein (1991) inverts the prevailing assumptions of much social science regarding citizen preferences. Economics has long treated consumer preferences as exogenously defined, and democratic theory views state policies as rightly arising from voter priorities.

In contrast, Sunstein (1991) argues that government may be *obliged* to override grass roots preferences, both in public policy and in the marketplace, and argues that such preferences should be regarded as strictly endogenous. In view of these diametrically opposed approaches, it would be a useful exercise to consider how such polar interpretations might best be validated.

In a second example, Gintis (2002) argues that, notwithstanding that it has informed decades of research, General Equilibrium Theory is flawed because its price mechanism is thoroughly exogenous and, therefore, unrepresentative of market interactions. Gintis introduces models designed to demonstrate the form that endogenous contract enforcement might take in labor, credit and consumer goods markets. To the extent that this argument is persuasive, it will result in a significant transformation of economic theory.

Each of the endogeneity examples is domain-specific. They illustrate the interaction of endogenous attributes as an inescapable aspect of social processes. In part, this is because, to capture their complexity, historical and policy-oriented issues must be represented in high-dimensional spaces. We now turn to such dimensional issues.

Social Dimensionalities. Consider historical questions of interest. What are the preconditions for revolution, and what can prior examples contribute to their understanding. What causes economic growth, and how can that potential be identified? Why do adjacent cultures converge, or resist convergence? In each case, there may be more specific issues that are dependent on these broader outcomes.

In the case of such questions, it may be beneficial to ask: what is the dimensional space in which the issue may be defined? Because of the breadth of questions, it takes little reflection to see that a dimensional space that could represent comparative cases is likely to be quite large. Indeed, the adequacy of the dimensional space may be considered a basic validation issue.

To start, there are multiple social science disciplines for a reason. In addition to the administrative efficiency they provide university administrations, they also constitute an implicit division of intellectual labor. Psychology, economics, political science, sociology and anthropology, each have the potential to contribute insight into a broad historical process. More specifically, each has concepts to which they commonly attend. Identity, market, state, social structure and culture are prototypical concepts from each of the listed disciplines, and it can also be recognized how these concepts might be relevant to revolution, economic growth, cultural convergence and a range of other policy-oriented topics.

However, these concepts are only prototypical and each discipline contains hundreds of competing or supplementary concepts that may need to be considered for a particular problem. Not

only does this significantly expand the prospective dimensional space for an historical issue, it also forces the analyst to face the inchoate nature of many social science concepts.

Quantitative social science recognizes the desirability of identifying orthogonal concepts, of course, but because most social science concepts exist in an (endogenous) semantic space, they are rarely that well-behaved. Efforts to impose dimensions using factor analysis or other data reduction techniques have not eliminated the problem. Accepting oblique dimensions acknowledges the problem, but the resulting "dimensions" are even less well behaved.

Different researchers understand the concepts in different ways, they evolve over time, and there are regional and/or disciplinary variations. These problems encourage modelers to accept or impose an arbitrary definition. While this is an understandable reaction, it does not directly confront underlying validation issues.

Trans-scale Propensities. Disciplinary distinctions have arisen, in part, from the need for cognitive economy. While this division of labor was probably inescapable when analysis and research was time intensive and personal, computational social science (CSS) provides a means for potentially reintegrating the social sciences. This enables the analyst to trace propensities across multiple scales on a scenario or case basis (cf., George & Bennett 2005:224-230).

Since any focal level may be influenced by lower or higher level social processes, CSS provides a strategy for representing cross-scale influences and, thereby, producing more integral social models. It has been argued that the most compelling processes to model will be those that operate at all levels, from the smallest to the largest, and the briefest to the enduring (Sallach 2010).

CONTEXT EFFECTS

In both the analytical and modeling communities, it is recognized that many social decisions are heavily influenced by the immersive context. While this characteristic is broadly applicable

within the social domain, it makes the validation process less deterministic and, therefore, much more difficult.

Here too, however, the general description (context effect) is too inclusive to be very useful. To better understand the implications for validation, it is important to grasp in greater depth the sources of social context effects.

Interaction, Paths and Strategies. One reason that context effects are inescapable in the social domain has to do with the role of interaction. From conversation to grand strategy, specific actor choices are dependent on the immediately prior communications, actions and/or events. Game theory captures interaction effects well, which is why it has become more prevalent, and even regarded as a possible basis on which to unify the social sciences (cf., Gintis 2009). Game theory itself is changing, being extended beyond previous limits (Scharpf 1997; Sallach 2000b; 2006; Vane & Lehner 2002), allowing interactive modeling to be applied to broader and more comprehensive issues.

One of the fundamental points about the significance of interaction is that it extends from the smallest to the largest social processes. During the twentieth century, sociologists made significant progress in documenting fine-grain interaction. Seminal contributions were made by Cooley (1907), Mead (1913), Schutz (1967; 1971), Goffman (1983), Garfinkel (1967; 2006), Heise (2006; Smith-Lovin 1988), Scheff (1990) and Rawls (1987; 1989).

As the foundational role of microinteraction has become more widely recognized, the hypothesis that interaction permeates institutions, up through the largest (Collins 1981; 2008; Kemper & Collins 1990; Stinchcombe 2001) is granted growing credibility. As Hilbert (1990:795) writes, "Interaction is scale-free, and occurs at all levels". One of the more coherent formulations of this theoretical framework is provided by Rawls (1989:166):

> Some aspects of both language and action have a sequential organization that is not derived from institutional constraints but is instead sensitive to needs of discourse which cut across

> social and institutional lines. However, every action and conversation takes place within an institutional context of some sort, and this context can always be brought to bear at the level of accounts.

This type of accounting or, as Bakhtin calls it, 'answerability' (cf., Kenny 2007), helps define the sequential structuring that weaves together trans-scale interactions.

However, while sequences of interaction are reasonably regarded as foundational, comparable processes arise in economic and state institutions. In recent years, economists have noted the way that small distinctions can grow into large-scale patterns. Described as 'path dependence', it explains how the QWERTY keyboard became established, and how computer standards and videotape formats become 'locked in' (David 1985; 1990), ultimately generating insight into the modern 'productivity paradox'.

Interaction creates strategic paradoxes as well. Luttwak (1987) points out that strategic choices are inherently paradoxical. The strongest most effective options are also the most obvious and, thus, are likely to be the best defended. Paradoxically, options that are suboptimal, and thus unexpected, may be the most effective.

Because it is open-ended, interaction is one source of social unpredictability. This open-endedness is mediated by interpretation. It is to this process that we now turn.

Interpretative Agency. In communication and action, human actors are oriented by meaning. Accordingly, in every situation, social actors consider, discern, define, attribute, convey, question, dispute, affirm, reconsider and evolve a meaning in a particular instance. Inevitably, the *attribution of meaning* is an indexical process: the same participants may view shared situations as having distinctive, or even conflicting, meanings. The process of meaning attribution is dynamic, often shifting rapidly, as the interpretation of the actor shapes and informs the subsequent flow of communications and acts. In all of these ways, the interpretive process contributes to context effects.

The modeling of orientation and meaning is not a new problem. Many significant strategies in artificial intelligence (AI), including semantic nets, logic-based semantics, rule-based inference, neural networks and subsumption, among others, have sought to address this capability. However, most AI implementations have limited sociality.

Models that advance the social sciences will, for many applications, need to generate the production, invocation, challenging and negotiation of meaning. More specifically, to capture the textured and situated nature of human action, we will need for models based on generic ontologies or prototypes to generate indexical methods and responses.

As Agre (1996:12) writes, "Perception and action … are inherently indexical in character." Indexical representations are tied, in epistemological and causal terms, to the agent's immediate circumstances, and are, thus, strongly endogenous to both model and actor.

Thus, while there are many exogenous CSS models that stipulate meaning, interpretive architectures are currently rare. As CSS moves forward, designs based on exogenous rules will increasingly be required to document that these rules provide determinate regularities, and have been validated as such. That is, if model outcomes are to be fully determined by external assumptions, model designers must expect to document that the possible social outcomes thereby generated are all that can be expected to arise in the corresponding domain.

Arguably, when pushed to the limits of modeling assumptions, action is controlled by meaning, exogenous factors or some kind of balance. The substantive significance of this choice is one topic that will need to be explored in the decade to come.

Inferring Intent. Whether individual or collective, human intent cannot be directly observed. Whether as researchers or as undistinguished participants in inchoate social life, we must infer the intent of our interlocutors from self-reports, past histories, social locations and more. But collective intentions remain the ground from which events unfold.

We attempt to distinguish earnestness from skilled deception, and settled intent from habit and drift. We recognize that declarative statements may not only be deliberately distorted, but may also be produced by one who lacks knowledge of future circumstances, or even of their own priorities. Our conclusions are often flawed or inaccurate. It is possible that intentions have a self-organizing fluidity and coherence that resist analytical representation (cf., Juarrero 1999; Roth 2000).

The problem of requiring data that cannot be directly observed is not unique to the social sciences (cf., Bell 2004), but it *is* pervasive. Inaccessibility of intent can be regarded as an ultimate source of social unpredictability. It ensures that social models lack an exogenous location from which to make validation assessments and, thus, must rely more heavily on inference based on behavioral mappings.

VALIDATING SOCIAL MODELS

The preceding overview has endeavored to provide a concise summary of some of the characteristics of the social domain that make model validation challenging, particularly from the standpoint of conventional validation strategies. However, if the case that validating social models is difficult has been adequately made, it may yet be unclear as to how the validation of social models can be accomplished. While this issue is inherently complex, and worthy of extensive discourse, it is possible to proffer a few precepts that may move such a discourse forward (Table 2).

The topics considered above, loosely characterized as: 1) non-linearity (malleable actors, causal fluidity, multivocality, sequence effects and correlated cases/contexts), 2) complexities (endogeneity, dimensional ambiguities, trans-scale propensities), and 3) context effects (interaction effects, endogenous interpretation, hidden intent) are recurrent and pervasive in social conjunctures of all types. Although not insurmountable, they cannot be ignored. The precepts are designed to address how social validation challenges can best be addressed.

The precepts presented in Table 2 are not intended as a full and definitive strategy for social model validation. On the contrary, they are a contribution to an emerging discourse on how to model social processes, and how best to validate those models. It is hoped that they will serve as a starting point for addressing social validation issues in greater depth and with greater sophistication.

Precept Number	**Precept Name**	**Precept Description**
1	Absence of determinacy	Social models cannot provide discrete, repeatable outcomes. Model results will always be expressed in probable paths and outcomes.
2	Interaction as foundational	Social phenomena are self-referential and evolve through interaction, which creates, *inter alia,* path dependencies and sequence effects.
3	Robust validation	The most effective validation strategy for social models is robust validation.
4	Challenge acknowledgement	The process of validating social models should address how non-linearities, social complexities and context effects are to be addressed, or why those categories of effects are not relevant to the model.
5	Substantive design	To the extent that social complexities can be addressed by theory integration and/or substantive social design, the validity of the resulting model will be enhanced thereby.

Table 2. Precepts of Social Validation

The contributions of Lempert, Popper and Bankes (2003) have made a unique contribution to the validation of policy-oriented models. Robust validation addresses both the reachability and stability of defined outcome spaces, and these are vital criteria in assessing social models.

However, current robust validation techniques were designed primarily with system dynamic models in mind. They need to be customized and extended in order to be applicable to social agent models and hybrid models that incorporate social agency. As robust validation techniques grow in maturity, social models validation issues will be able to be addressed more effectively.

CONCLUSION

The sources of social unpredictability discussed in this paper are neither an elegant synthesis nor exhaustive. Rather, it is an illustrative tour of the many sources of complexity and unpredictability. These sources are interleaved and coevolving. In some settings, they reinforce each other; in others, they compensate for each other. In either case, an effective validation strategy for social models must acknowledge them.

The validation of social models is challenging, but not impossible. Effective strategies are as yet embryonic, and their further evolution is an essential aspect of the effective use of social models. However, there cannot be hope for effective social validation *until* the unique issues associated with social science validation are recognized, conceptualized and addressed.

Such recognition is the prerequisite for a shared validation discourse that can: 1) identify fundamental challenges, 2) define and constrain the foci of social validation, and 3) suggest strategies that can resolve or mitigate emerging difficulties. Such advances are required in order to achieve practical benefits from ever more effective social models. It is hoped and intended that, by framing these issues in greater depth, this discussion contributes to progress in the effective use of social models.

REFERENCES

Abbott, Andrew. 1988. "Transcending general linear reality." *Sociological Theory* 6:169-186.

—. 2001. *Time Matters: On Theory and Method*. Chicago: University of Chicago Press.

Agre, Philip S. 1996. Computational research on interaction and agency. Pp. 1-52 in P.E. Agre and S.J. Rosenschein, eds., *Computational Theories of Interaction and Agency*. Cambridge, MA: MIT Press.

Bell, John S. 2004. The theory of local beables. Pp. 52-62 in *Speakable and Unspeakable in Quantum Mechanics*. New York: Cambridge University Press.

Burns, Tom R. & Helena Flam. 1987. The Shaping of Social Organization: Social Rule System Theory with Applications. London: Sage Publications.

Collins, Randall. 1981. "On the microfoundations of macrosociology." *American Journal of Sociology* 86:984-1014.

—. 2008. *Violence: A Micro-Sociological Theory*. Princeton, NJ: Princeton University Press.

Cooley, Charles H. 1907. "Social consciousness." *American Journal of Sociology* 12:675-694.

David, Paul A. 1985. "Clio and the Economics of QWERTY." *American Economic Review* 75:332-337.

—. 1990. "The dynamo and the computer: An historical perspective on the modern productivity paradox." *American Economic Review* 80:355-361.

Garfinkel, Harold. 1967. *Studies in Ethnomethodology*. New York: Prentice-Hall.

—. 2006. *Seeing Sociologically: The Routine Grounds of Social Action*. Boulder, CO: Paradigm Publishers.

George, Alexander A. and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.

Gintis, Herbert. 2002. "Some implications of endogenous contract enforcement for general equilibrium theory." Pp. 176-205 in *General Equilibrium: Problems and Prospects*, edited by F. Petri and F. Hahn. London: Routledge.

—. 2009. *The Bounds of Reason: Game Theory and the Unification of the Social Sciences*. Princeton, NJ: Princeton University Press.

Hartley, Dean and Stuart Starr. 2010. Verification and validation. Pp. 311-336 in A. Kott & G. Citrenbaum, eds., *Estimating Impact: A Handbook of Computational Methods and Models for Anticipating Economic, Social, and Security Effects in International Interventions*. New York: Springer.

Heckelman, Jac C. and Hakan Berument. 1998. "Political business cycles and endogenous elections." *Southern Economic Journal* 64:987-1000.

Heise, David R. 1979. *Understanding Events: Affect and the Construction of Social Action*. New York: Cambridge University Press.

—. 2006. "Sentiment formation in social interaction." Pp. 189-211 in *Purpose, Meaning and Action: Control Systems Theories in Sociology*, edited by K. A. McClelland and T. J. Fararo. New York: Palgrave Macmillan.

Heritage, John. 1984. *Garfinkel and Ethnomethodology*. Cambridge, MA: Polity Press.

Hilbert, Richard A. 1990. "Ethnomethodology and the micro-macro order." *American Sociological Review* 55:794-808.

Jacoby, William G. 1991. *Data Theory and Dimensional Analysis*. Newbury Park, CA: Sage Publications.

Juarrero, Alicia. 1999. *Dynamics in Action: Intentional Behavior as a Complex System*. Cambridge, MA: MIT Press.

Kemper, Theodore D. and Randall Collins. 1990. "Dimensions of microinteraction." *American Journal of Sociology* 96:32-68.

Kurihara, Kenneth K. 1960. "An endogenous model of cyclical growth." *Oxford Economic Papers* 12:243-248.

Lempert, Robert J., Steven W. Popper, and Steven C. Bankes. 2003. *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis*. Santa Monica, CA: RAND.

Luttwak, Edward N. 1987. *Strategy: The Logic of War and Peace*. Cambridge, MA: Harvard University Press.

Kenny, Robert Wade. 2007. "The good, the bad and the social: On living as an answerable agent." *Sociological Theory* 25:268-291.

Mead, George Herbert. 1913. "The social self." *Journal of Philosophy, Psychology and Scientific Methods* 10:374-380.

Rawls, Anne Warfield. 1987. "The interaction order *sui generis:* Goffman's contribution to social theory." *Sociological Theory* 5:136-149.

—. 1989. "Language, self and social order: A reformulation of Goffman and Sacks." *Human Studies* 12:147-172.

Roth, Abraham Sesshu. 2000. "The self-referentiality of intentions." *Philosophical Studies* 97:11-52.

Sallach, David L. 2000a. "Classical social processes: Attractor and computational models." *Journal of Mathematical Sociology* 24:245-272.

Sallach, David L. 2000b. "Games social agents play: A complex form." in *Joint Conference on Mathematical Sociology in Japan and America*. Honolulu.

—. 2006. "Complex multigames: Toward an ecology of information artifacts." Pp. 185-190 in *Proceedings of the Agent 2006 Conference on Social Agents: Results and Prospects*, edited by D. L. Sallach, C. M. Macal, and M. J. North. Chicago: Argonne National Laboratory.

—. 2010. "Toward a social algebra." in *the Meetings of the American Sociological Association*. Atlanta.

Scharpf, Fritz W. 1997. *Games Real Actors Play*. Boulder, CO: Westview Press.

Scheff, Thomas J. 1990. *Microsociology: Discourse, Emotion and Social Structure*. Chicago: University of Chicago Press.

Schutz, Alfred. 1967. *The Phenomenology of the Social World*. Translated by G. Walsh and F. Lehnert. Evanston, IL: Northwestern University Press.

—. 1971. "Choosing among projects of action." Pp. 67-96 in *Collective Papers I: The Problem of Social Reality*, edited by M. Natanson. The Hague: Martinus Nijhoff.

Stinchcombe, Arthur L. 2001. *When Formality Works: Authority and Abstraction in Law and Organizations*. Chicago: University of Chicago Press.

Sunstein, Cass R. 1991. "Preferences and politics." *Philosophy and Public Affairs* 20:3-34.

Vane, Russell and Paul Lehner. 2002. "Using hypergames to increase planned payoff and reduce risk." Pp. 315-336 in *Game Theory and Decision Theory in Agent-Based Systems*, edited by S. Parsons, P. Gmytrasiewicz, and M. Wooldridge. Boston: Kluwer Academic Publishers.

Wlezien, Christopher, Mark Franklin, and Daniel Twiggs. 1997. "Economic perceptions and vote choice: Disentangling the endogeneity." *Political Behavior* 19:7-17.

Jean Scholtz, A User-Centered Approach to Social Modeling and Simulation for Decision-making

Presented at the DTRA/Sandia Workshop on Challenges in Computational Social Science. Santa Fe, New Mexico. October 25-28, 2010.

Abstract

The goal of much software research is to incorporate the research into a larger software package for a set of end-users. In many cases, users are not engaged or consulted in the process or if they are it is not until the development phase. This paper presents an argument for involving users much sooner in the research and development process. In early user involvement we are investigating the utility of the software. While software usability is certainly necessary for end-users to be able to work effectively, it is not sufficient for end-user adoption. In order to convince end-users to adopt new software it is necessary that they perceive some benefit. Understanding metrics that are important to end–users and measuring the impact that software use can have on those metrics is one way to achieve this. While it is advisable to follow this process for software research being incorporated into applications that end-users will consume, it is essential to follow this process when the software is being used for high impact decision-making. Human social, cultural and behavioral modeling and simulation is an area that will certainly benefit from a user-centered approach. Indeed, the feasibility of incorporating these models into decision-making without having considered user tasks and concerns in the R&D process seems highly unlikely. Recommendations for a user-centered approach to human social, cultural and behavioral modeling are discussed. These recommendations are from existing literature and author's experience with different research projects, including an effort in techno-social predictive analytics [1].

1 Introduction: Usability versus Utility

A user-centered design approach takes into account both usability and utility concerns. The International Organization for Standardization (ISO) defines usability as "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use [2]." Utility is a broader assessment in that it concerns the overall process and work flow of the targeted end-users and stakeholders and how the new product will fit into that process. Utility metrics, however, are not generic but are customized for the targeted users. Utility evaluations necessitate working with various end-users and stakeholders to identify metrics that reflect utility [3]. These measures are unique to the values of the organization and to the various stakeholders and end-users. Potential metrics might be a better quality product, a safer or more efficient process, or more confidence in the analysis. These metrics need to be agreed upon by researchers, end-users and evaluators, and used for utility assessments throughout the R&D cycle. If utility evaluations during the cycle do not show marked improvements, then researchers need to be able to assess the problems and take the research in a direction that will produce the needed improvements. While usability evaluations ensure that end-users can effectively use software, utility evaluations ensure that there is a measurable benefit in doing so. An advantage of utility evaluations is that end-users can easily understand the benefits of adopting the new tools. If those benefits are such that they outweigh the disruption and cost of change, then the new technology should be more readily adopted.

Although we tend to use the term "utility evaluations", it should be clarified that we use both formative and summative evaluations and that in the process of identifying the metrics for utility, we are essentially establishing high level user requirements that need to be considered in providing the new technological capabilities. Usability concerns need to be addressed when the architecture for the system is being developed. Decisions about which software modules talk to which other modules, what tasks are being automated, what provisions for backtracking are included all have implications for usability of the software [4].

Utility decisions are important to consider in earlier stages. When users are being provided software tools that provide new capabilities, we need to understand the impact of the technology on their current processes and products.

2 The User-Centered Approach to Technology Research

Mayhew [5] outlines three phases of a usability engineering lifecycle: requirements analysis, design/testing/development, and installation. The requirements phase involves doing a user task analysis, developing user profiles, understanding basic platform capabilities and constraints, and producing a set of usability goals. These usability goals are commonly expressed in terms of times for users to achieve a task, number of errors in a benchmark task, user satisfaction ratings for ease of use. In this lifecycle, the purpose of the user task analysis and the user profiles is to understand how the user will interact with a specific product that has already been defined. In taking a user-centered approach to research in software capabilities, we are more interested in understanding what functionality particular users need and the impact this could have on their work.

Chen and Atlee [6] note that requirements engineering is about defining precisely the problem that the software is to solve. They note 5 steps in requirements engineering: elicitation, modeling, requirements analysis, validation and verification, and requirements management. The elicitation step can range from looking at modifications to an existing software application to determining a new product based on new software research.

These software engineering processes were developed for the process of producing specific applications from existing technology. In applying the user-centered approach to research-based software, we have a slightly different approach for requirements engineering. We actually have two communities, the researchers and the potential end users, to bring together to determine if the research can be useful to the potential end users. To do this, we need to understand the end-users current work, their current problems, their value system, and any constraints that the work environment, processes, customer interaction, etc. imposes on them. It is important to understand that this is a holistic approach. In order to understand the impacts of the specific research, we need to understand, not only the particular aspect of the users' problem that could potentially be eliminated, we need to understand the entire flow of work to make sure that no indirect affects occur. We also need to understand the technology research and the potential benefits that can be realized.

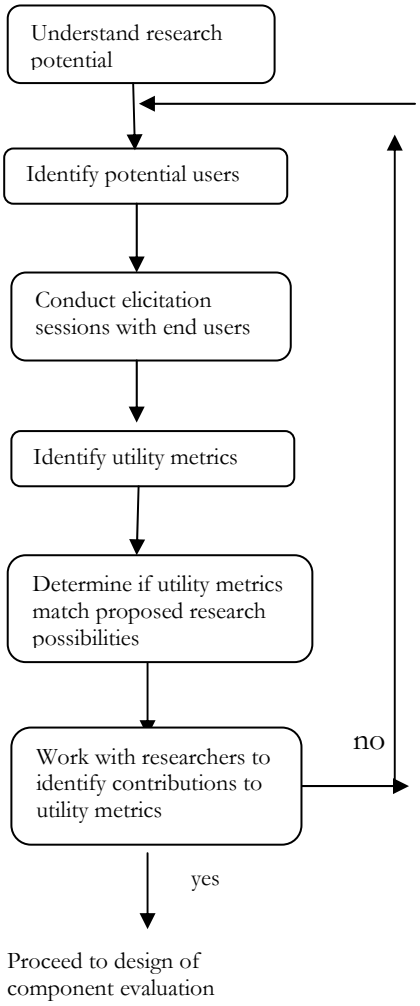


Figure 1. The first step in user involvement in the research process

We agree with both Mayhew's approach to usability engineering and the requirements tasks in requirements engineering. Our focus, however, is on bringing users into the research process. This involves a slightly different process as the technology may go well beyond what end-users currently have. Figure 1 shows the steps in our approach. This is a compromise between the push and pull of technology. Our goal is to determine if and how the technology being developed in a particular research project will provide utility to a specific group of end users. The first task is identifying potential end users. These might already be identified in a government funded research project or in an industry effort. If this is not the case, then we will have to brainstorm with the research community and identify contacts in different user groups. If this is the case, then an initial but less

intensive round of elicitation with several groups should be done to find the best possibility. For more intensive elicitation once a promising end-user population has been identified, techniques such as task analysis, cognitive task analysis [7], ethnographic techniques, and other interview techniques and observations [8] should be employed.

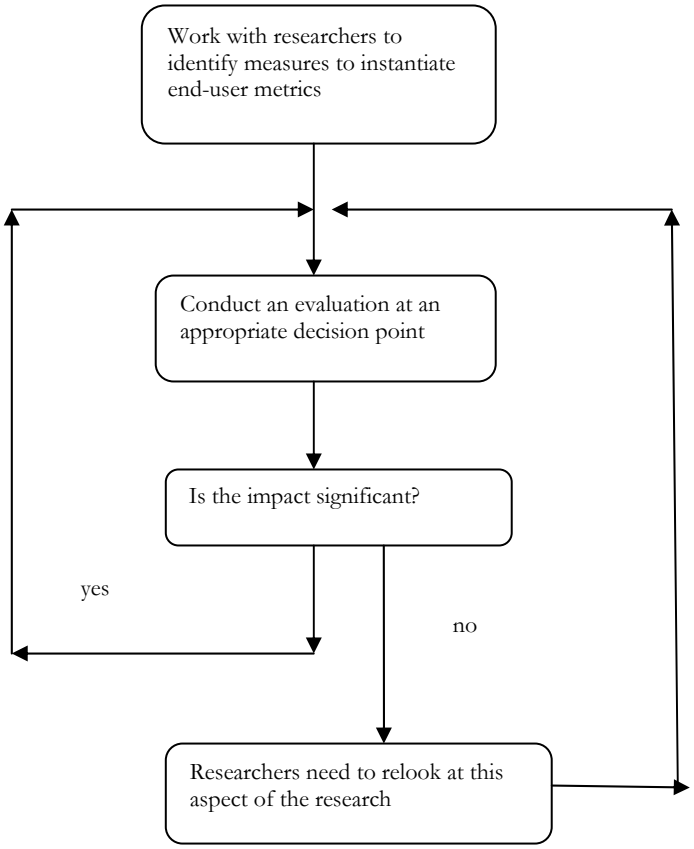


Figure 2. The second step in user involvement in the research process

An important addition to the normal techniques is to determine the utility metrics that are important to the end user population. Normally these metrics are such things as time, quality, or production capability. For example, in an interview with a number of Port Police who are on duty at a major airport, two metrics were identified: time to respond to an incident and time spent on the airport floor. Time to respond could be improved by functionality that helped officers know when other officers were involved in investigation work and help the others move to strategic locations to lessen response time for other incidents. The police staff felt that their visibility in the airport deterred crime. So any time that they had to leave the airport proper and go to their offices for forms, filing reports, and such took away from this visibility. Providing functionality that allowed the officers to conduct some of their paperwork remotely would lessen the time they needed to spend off the airport floor. During the elicitation process it is necessary to determine what is important to the user community and use those metrics to measure the impact of the new technology.

The next step is to work with the researchers to determine if their technology can positively impact any of these measures. If this is not feasible, then we need to start looking for another possible home for the technology. But if it is, then we need to work further with the research community to determine what the impact is and how we can measure it.

The goal is to come out of the research phase with an identified end-user population, a set of utility metrics that have been identified for the end user and technology that has been shown to be useful to this end user population through the research phase. Figure 2 shows how the research phase uses these metrics to focus aspects of the research. The evaluations are conducted with either members of the end user community or surrogates for these end-users. The surrogates are selected to match specific skills of the end users where the skills depend on the particular evaluation being conducted.

The evaluations should be done to provide researchers with an answer to a specific question in the research and should be done as soon as possible. User evaluations do not necessitate finished or even working software. Paper prototyping [9] and Wizard of OZ [10] methods can provide answers to many questions. However, determining methods for these different types of evaluations is an exercise in creativity.

It is helpful to determine a benchmark once the end user community has been identified. For example, it the end users are intelligence analysts, it might be helpful to give a group of analysts a specific task and a collection of documents. Recording the documents they used, the time it took them in various phases of analysis, any particular analytic techniques they used, the processes used, the products produced and the ranking of those products based on quality criteria would all be useful in comparing the impact of any technologies designed to help intelligence analysts in their work [3,11,12]. The task, of course, needs to embody the problem that the technology is helping to overcome.

3 Applying this to Human Social, Cultural and Behavior Modeling and Simulation Research

McNamara and Trucano [13] have laid out ten challenges that need to be addressed in order to develop decision support technologies and tools in computational social science modeling and simulation for national security decision making. The challenges for human-computer interaction are:

- how to transform specialized modeling and simulation prototype technologies into tools for non-expert users;
- how to evaluate how these tools influence the user community's performance in exploratory problem solving and sensemaking; and
- how to create adoptable computational modeling and simulation technologies

These challenges are not independent; rather the modeling and simulation technologies will be adopted if they are created so that non-expert users can use them and if evaluations show that analysts and decision makers can make "better" decisions using the tools than without the tools. In the following sections we discuss how we can use information from our user-centered approach to technology research to provide insights to address these challenges.

3.1 End-User Metrics

If we assume that we have identified the intelligence community as a group of potential end-users, we have 2 subgroups to satisfy: the analysts who will be working with the models to determine different possibilities and who will need to communicate their recommendations and the decision-makers who will be using the recommendations in conjunction with other information to make policies. Some studies have already been done that shed some light on how analysts work with technosocial modeling [14,15,16].

Pfautz et al. [14] analyzed the application of modeling techniques and technologies based on Cognitive Systems Engineering. Cognitive Systems Engineering analyses users engaged in work and uses an iterative design, implementation and evaluation cycle. In particular they analyzed user requirements in the process of developing a number of applications based on the user of models and found that analysts are very concerned with the validity of the models. In some instances analysts felt it was necessary for the models to be created within their community in order to be trusted. The bottom line is that if analysts do not trust a model, they will not use it. The recommendations analysts make will be based in part on the information from the model. And, as with any other source used, a model must be reliable and trusted. Therefore, Pfautz et al. are focused on building tools to help analysts create their own models, including building templates of models that end-users can customize.

Hanson and Russell [15] note that metrics for technosocial predictive analytics must satisfy two audiences: those testing the model's theoretical basis and the operational user who has to have a reason to trust the model. They suggest that validation is easier to accomplish, and therefore easier for the user to trust the model, if the model is approached as a black box. That is, if the inputs and outputs from the black box correspond to input and outputs in the real world, it is not necessary that users understand that specific function implemented in the model.

Klein et al. [16] found that decision- makers need not only situation awareness, the facts about the environment or event that is occurring, but also need to understand the options available to them and the consequences of selecting each action. Decision-making has several constraints: the time in which the decision has to be made and the fidelity of the information that is needed to make the decision. Klein et al. conducted a study to look at the essential precision and fidelity of models needed to support decision spaces. They found an example where a lower fidelity model significantly changed the decision space. They conclude that more work needs to be done in looking a boundary conditions and determining where models are unable to provide decision-quality data.

These studies have identified potential end-user metrics. For analysts, trust in the model is critical. For decision-makers, we need to consider the time frame in which the decision is needed and ensure that analysts working with the software can produce the recommendation within that time frame. We need to have a measure of the fidelity of the information needed for making the decision and ensure that analysts can deliver the options available with the necessary fidelity. Interviews with specific groups of end users will produce additional metrics and more specific values for trust, time constraints and information fidelity.

3.2 How to Make Modeling Tools Available to Non-Expert Users

In the above discussion, we identified trust as a measure of importance for the analysts. The question then is not just how to make modeling tools available to non-expert users but how to make these tools available and ensure that analysts trust the information from the modeling tools?

Mostashari [17] advocates a collaborative modeling approach with stakeholders. He notes that a modeling process that is being used to support decision–making should produce more than a system that meets technical requirements. Dürrenberger et al. [18] notes that good models for science-intensive decision-making should:

- have a high degree of visualization and interaction
- have simple, transparent structures and produce results quickly
- should include links to local tangible issues
- should not be regarded as the sole source of information

Mostashari notes that technical engineering models can improve communications in planning large-scale engineering systems. In fact, a systems model should facilitate communications among representatives from different domains and perspectives. To facilitate this Mostashari and Sussman developed a stakeholder-Assisted Modeling and Policy Design Process [19]. The take-away from this effort is that the use of modeling and simulations would be advanced by having stakeholders participate in the design of the actual system.

While participatory design is a technique often used in the human-computer interaction field, the issue is finding analysts and decision-makers who have the time needed to participate in these activities. Other possibilities include using historical data along with the modeling and simulation tools and comparing the decisions to those of the historical decision. This is problematic as the existing data may not be sufficient for the models and the participants in such a study should not be aware of the situation and its outcome. If a different decision is made than the historical decision, we are also faced with trying to determine if this outcome would have been "better."

Another possibility is to engage representatives from the end-user community in user-centered evaluations during the creation of modeling tools. This technique was the approach taken in the Technosocial Predictive Analytics Initiative [http://predictiveanalytics.pnl.gov/about.stm] at the Pacific Northwest National Laboratory. The concept was to produce a simplified visual representation of the model showing the various parameters that served as inputs and outputs and a simplified representation of other parameters used within the model. The methodology proposed was to use this visual representation and have the analysts study it and explain to us what the model did. Then we would ask the analyst to walk through a scenario and "use" the model to obtain information pertinent to the scenario. This information would be compared to the information gathered in the experiment conducted with domain experts. The differences would be analyzed to determine where differences occurred and that would be used to adjust the fidelity of the representation if necessary. Table 1 shows the questions to be addressed using this technique.

Evaluation component	Model
Question to be addressed	Is the model understandable? Are the inputs correct? Are the outputs correct? Are the users able to understand the impacts of parameters changes on other parts of the models? What additional scenarios could be explored?
Target users	Domain experts; senior analysts
Materials	Paper based schematics/ descriptions of models
Method for collecting data	Open ended interviews Think aloud exploration of a scenario

Table 1. A methodology for a user evaluation of a technosocial model

This could be used in determining what information end-users need to be able to understand and trust the information that a model provides. If a representation alone does not provide engender trust in the model, then the next step would be to determine if the addition of some actual computations for several different scenarios would suffice.

3.2.1 Making the Information Available to Decision-makers

The next step is to understand how decision-makers utilize information provided to them from computational social models. Questions we need to answer are what information do decision-makers want? At what fidelity is this information needed? How can information from models be presented and how much reliance on this information will decision-makers feel comfortable making?

There have been numerous studies on how people make decisions. A number of these have taken place in the operational world. Naturalistic Decision Making [20] studies show that people rarely compare options when making a decision. Rather they use their experience to size up the situation, classify it and identify a typical response to that particular type of situation. Rather than attempting to find the optimal solution, they are content to find a satisficing solution. This may also be attributed to the strict time constraints found in operational environments. For operational situations, it has been suggested that the best aid would be to provide tools that would provide with a better understanding of the situation. Once the situation is understand, the appropriate course of action is usually obvious.

In operational decision-making, mental simulation is used for evaluating a course of action. This is a means of doing a deep search on a few options, in the context of the situation, but not as a means for comparing multiple possibilities. In comparing multiple options all strategies used basically employed breaking the course of action into component and comparing these [21].

Etzioni [22] presents three approaches to social decision making. Rationalistic models assume that the decision-maker becomes aware of the problem, determines a goal, carefully weights the approaches to achieve this and selects one based on estimates of the pluses and minuses of each. A

problem with the rationalistic approach is an inability to collect and utilize all of the information in a timely fashion. Additionally, weighing the merits of each approach implies that the decision-maker has a set of values against which these can be evaluated. The incrementalist approach is the other extreme. The decision–maker focuses only on those alternatives which differ incrementally from existing policies. For the alternatives to be considered, only a few important consequences are evaluated. There is no right decision, but a number of decisions that are made incrementally. Radical change does not come about using the incremental approach.

Etzoni proposes a third approach to decision-making, a mixed-scanning approach. In this strategy a detailed examination of some aspects of the situation are made, with a scanning done of the other aspects to pick up any unexpected behaviors. Part of the strategy is to decide which aspects deserve a more detailed look and which are suitable to scan. This allocation can, of course, change over time if the environment is radically changing. If we determine that decision-makers are indeed comfortable a particular approach, this will impact what the model and the fidelity of the information needed.

Another question is how to provide information from models to decision-makers. The field of visual analytics is based on the premise that interactive visualizations [23] help information analysts better understand large amount of information; thus increasing their situation understanding. While researchers in the field are still exploring metrics to show the utility of visual analytic tools [23], the analysts beginning to use these systems present their analysis to decision-makers in text. A challenge for researchers is to provide analysts systems that help them present their analysis visually [25]. This still leaves us with the problem of determining what type of presentation is better?

It has been suggested that if a "better" decision is made than the presentation of information may have been more effective. But what is a "better" decision? In the situation awareness literature [26] having better situation awareness does not necessarily mean that a better decision is made or that the end result is better. A decision-maker can have perfect situation awareness and make the wrong decision for various reasons. And by pure luck a decision –maker with imperfect situation awareness, could select a very reasonable course of action.

Our proposal is to again use the metrics important to the decision-maker to evaluate the effectiveness of the decision. Although, we speculate that in many cases, these metrics may be specific to a certain domain at a certain time. For example, the goals of a CEO in difficult economic times may be radically different those goals in times of prosperity.

As many of today's complex problems require expertise from many different domains, a further question is whether models substitute for domain expertise? Again, this can be measured in terms of trust.

An interdisciplinary teams of researchers is need to devise user studies to help understand the issues that face analysts and decision-makers using computational social models.

3.3 How to Evaluate How these Tools Influence the User Community's Performance in Exploratory Problem Solving and Sense-making?

Pirolli and Card [27] studied the sense-making loop of intelligence analysts and proposed a leverage points where technology could have a decided impact. They identified these leverage points within the two major loops in the intelligence analyst's process: foraging and sense-making. In foraging, analysts are trying to identify the relevant material from a large mass of existing documents. There is a cost associated with exploring more material to make sure more relevant material is there to possibly exploit. Scanning for information, shifting attention to other domains for information, doing follow-up searches consume resources. More information also means more time needed to analyze and exploit that information.

The sense-making look is affected by many issues noted by Heuer [28]: generating alternative hypotheses, confirmation bias, and the span of human attention to evidence and alternative hypotheses.

To evaluate how modeling tools can impact the end-user's problem solving and sense-making processes it is feasible to use both quantitative and qualitative measures. Examples from several researcher projects [3,11,12,29] looked at measures such as time spent in different phases, number of documents examined, quality of the report generated (as ranked by participants), hypotheses generated, and participants confidence in their recommendations.

3.4 How to Create Adoptable Computational Modeling and Simulation Technologies?

As we stated in earlier, analysts will only use tools that they trust. So the first step is to understand what will provide the necessary information for analysts to trust a given model. Does the model need to have a proven track record? Can it be a black box model whose inputs and outputs are in line with historical events in the real world? Do analysts want to understand the parameters and relationships used in the model? Does the model provide analysts with information that the decision-makers will use?

Finding methodologies to do user-centered evaluations and using these to instantiate the metrics obtained from the end-users will go a long ways to convincing the user community that they will benefit from adopting these technologies [30].

4 Conclusions

Today's analysts and decision makers are extremely busy. They are inundated with people who have great technologies to help them. To get their attention it is necessary to convince them of the utility of tools. One way to do this is for them to see others successfully using the tools [31]. But this takes a long period of time especially for organizations not known as early adopters. Another approach is to provide them with measures of utility in terms they understand: their own metrics for success in their jobs. Determining those and then instantiating them from measures of the technologies can show the impact that the technology will have in a real world environment.

There are challenges to doing this. Technology researchers need access to the end-users to determine those metrics. Creating the necessary user-centered evaluations is not anywhere as straight forward as doing a usability test. The researchers need sample user tasks and sample data. They need independent evaluators to conduct the evaluations. Not only do the evaluations need to

produce measures but it is necessary to produce insights into why the numbers turned out as they did in order for the researchers to refine their work.

A community effort is needed to make this work feasible. We need both research efforts and practical efforts. Interdisciplinary research is needed to understand the issues facing analysts and decision-makers using computational social models. As many researchers are already building such modeling and simulation capabilities, user-centered evaluation techniques can be used to help refine these tools. The initial evaluations will be difficult but sharing the lessons learned with the community will ensure future improvements in the evaluation process and the technology research and will lead to a facilitated transition to the end –users.

References

J. Scholtz and M. Whiting, "User-Centered Evaluation of Technosocial Predictive Analysis, " *AAAI Spring Symposium, Technosocial Report SS-09-09*. AAAI Press: Menlo Park, CA. pp. 126-129. 2009.

ISO 9241-11. Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) Part 11: Guidance on Usability. ISO. 1998.

J. Scholtz, "methods for Evaluation Human Information Interaction Systems," *Interacting with Computers,* Vol 18(4), July, pp. 507-527. 2006.

L. Bass and B. John, "Achieving usability through software architectural styles," *CHI'00 extended abstracts on Human Factors in Computing Systems,* The Hague, The Netherlands, pp. 171-172. 2000.

Mayhew. *The Usability Engineering Lifecycle",* London: Academic Press. 1999.

B. Cheng and J. Atlee, "Research Directions in Requirements Engineering", *FOSE '07: 2007 Future of Software Engineering.* May, 2007.

E.M. Roth and R. Mumaw, "Using cognitive Task Analysis to Define Human interface Requirements for First-of-A-kind Systems," *Human Factors and Ergonomics Society Annual Meeting proceedings 39th, Oct,. 9 – 13,* San Diego, CA. 1995. Pp. 520-524.

T. Salvador and J. C Scholtz, "Systematic creativity: a methodology for intergrating user, market and engineering requirements for product definition, design and usability testing," *Proceedings of the IFIP TC2/WG2.7 Working Conference on Engineering for Human-Computer Interaction. London:* Chapman & Hall, Ltd. August, 1995.

Beyer and K. Holtzblatt, "Contextual Design", *Interactions,* Vol. 6 (1). New York, NY: ACM. Jan/Feb 1999. Pp. 32-43.

N. Dahlbäck, A. Jönsson, and L. Ahrenberg, "Wizard of Oz Studies – Why and How", *Knowledge Based Systems,* Vol. 6(4), 1993.

Scholtz, E. Morse, and T. Hewett, "In Depth Observational Studies of Professional Intelligence Analysts 2005," in *Proceedings of the International Conference on Intelligence Analysis* MacLean, VA: Available at https://analysis.mitre.org, accessed 19 June 2010, 2006.

J. Scholtz, E. Morse, and M. P. Steves, "Evaluation Metrics and Methodologies for User-Centered Evaluation of Intelligent Systems," *Interacting with Computers,* pp. 1186-1214, 2006.

McNamara, L. and Trucano, DTRA Challenges Workshop. 2010.

J. Pfautz, E. Carlsen, D. Koelle and E. Roth, "User-Created and User-Adaptable Technosocial Modeling Methods', *AAAI Spring Symposium, Technosocial Report SS-09-09*. AAAI Press: Menlo Park, CA. pp. 88-93. 2009.

L. Hanson and A. Russell, "The Challenge of the Black Box: Technosocial Predictive Analytics (TPA) and the Real World", *AAAI Spring Symposium, Technosocial Report SS-09-09*. AAAI Press: Menlo Park, CA. pp. 55-58. 2009.
G. Klein, M. Pfaff, and Drury, J., "Supporting a Robust Decision Space", *AAAI Spring Symposium, Technosocial Report SS-09-09*. AAAI Press: Menlo Park, CA. pp. 88-93. 2009.pp. 66-71.

Mostashari. "Collaborative Systems Modeling for Complex Infrastructure: The Case of the Cape and Islands Collaborative Regional Energy Planning", *AAAI Spring Symposium, Technosocial Report SS-09-09*. AAAI Press: Menlo Park, CA. pp. 82-87. 2009.

G. Dürrenberge, H. Kastenholz, and J. Behringer. Integrated assessment focus groups: bridging the gap between science and policy? Science and Public Policy, Volume 26, Number 5, 1 October 1999 , pp. 341-349(9)

Mostashari and J. Sussman. Stakeholder Assisted Modeling and Policy Design for Environmental Decision-making. Journal of Environmental Assessment, Policy and Management, September 2005.

G. Klein. Naturalistic Decision Making: Implications for Design. DTIC Report CSER1AC SOAR 93-1. April 1993.

C.E. Zsambak, G. Klein, M. N. Kyne and D.W. Klinger. Advanced Team Decision Making: A developmental model (prepared for the US Army Research Institute for the Behavioral and Social Sciences). DTIC NO.. AD-A259512.

Etzioni. Mised-Scanning: A "Third" Approach to Decision-making. Public Administration Review, Vol. 27 (5). Pp. 385-392. 1967.

J. Thomas and K. Cook, Illuminating the Path. Los Alamitos, CA: IEEE Computer Society. 2005.

C. Plaisant, G. Grinstein, and J. Scholtz, "Visual-Analytics Evaluation," *IEEE Computer Graphics and Applications,* vol. 29, pp. 16-17, 2009.

N. Chinchor, N. and W. Pike, W. The science of analytic reporting, Information Visualization, 8: 286-293. doi:10.1057/ivs.2009.21. 2009.

Endsley, M., Bolté, B, and Jones, D. (2003) Designing for situation Awareness, New York, NY: Taylor & Francis.

Pirolli, P. and Card, S. (2005) The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through cognitive Task Analysis. In Proceedings of 2005 International Conference on Intelligence Analysis.

R. J. Heuer, Psychology of Intelligence Analysis. Center for the Study of Intelligence. Central Intelligence Agency. 1999.

D. Kelly, P.B. Kantor, E.L. Morse, J. Scholtz and Y. Sun. Questionnaires for eliciting evaluation data from users of interactive question answering systems, Natural Language Engineering 15 (1):

119–141. c_ 2008 Cambridge University Press doi:10.1017/S1351324908004932 Printed in the United Kingdom

Scholtz, J., Cook, K., Whiting, M.A., Lemon, D. and Greenblatt, H. (2009) Visual Analytics Technology Transition Progress, Information Visualization, 8:294-301. 2009. Doi: 10.1057/ivs.2009.20.

G. Moore. Crossing the Chasm. New York, NY: Harper Collins. 2002.

Jessica Glicken Turnley, Assessing the 'Goodness' of Computational Social Models

Presented at the DTRA/Sandia Workshop on Challenges in Computational Social Science. Santa Fe, New Mexico. October 25-28, 2010.

Acknowledgements: I would particularly like to acknowledge the contribution of Aaron Perls of Galisteo Consulting Group to this paper. He has given me good ideas and much background material to support and focus my own ill-formed thoughts. I also would like to thank Jennifer Perry of DTRA/ASCO for her long-time support of my interest in this area, Laura McNamara for organizing this symposium and inviting me to it, and all my colleagues who have listened and responded to my soapbox speeches on this topic over the years.

Introduction

Computational representations of social phenomena (computational social models) have become increasingly visible in a variety of planning and analysis environments, including corporate, environmental and military/national security arenas. The increased visibility and the increasing sophistication of the models themselves have raised a variety of methodological questions.

Computational models of physical and biological phenomena have been around for almost as long as computing. Historically, the assessment of goodness of computational models has been phrased as a validation question, more specifically as a question of predictive validation. This arises from a fairly constrained notion of what these types of models are designed to do – i.e. to predict. As we shall see later in this discussion, models (of which computational models are just a subset, and computational social models an even smaller subset) can serve a variety of purposes. I shall argue that the inherent nature of computational social models poorly suits them for prediction. If we adjust our notion of the purpose of these types of models, we will find a corresponding adjustment in the mechanisms by which we assess how well they serve that purpose. While predictive validation may still have a role, it will be a greatly diminished role in favor of other types of assessment, including other types of validation.

Definitions

Any discussion of validation must begin by clearly distinguishing validation from verification. The two processes are often presented in tandem or even treated as an integrated process, as in a 'V&V [verification and validation] capability.' However, they require significantly different skills and fill very different functions.

Verification

The IEEE handbook on software engineering terminology defines verification as "formal proof of program correctness,"[63] where 'formal proof of program correctness' means "a formal technique used to prove mathematically that a computer program satisfies its specified requirements."[64] As the U.S. Federal Drug Administration guidance for software developers puts it (in conformance with ISO 9000), "Software verification looks for consistency, completeness, and correctness of the

[63] IEEE Software Engineering Terminology IEEE-STD 610.12-1990 p.81

[64] Ibid. p.59

software."[65] The U.S. Department of Defense, an historic developer and user of physical models (and an agency with a recent interest in computational social models), defines verification as "the process of determining that a model or simulation implementation accurately represents the developer's conceptual description and specifications."[66] Verification thus is a determination of internal consistency– that if the programming specifications required that the code does certain things, it did. Various tools have been developed to determine the 'goodness of fit' of the code performance to specifications, usually through statistical measures. Verification is not designed to challenge the specifications: those are taken as given.

Validation

Validation is much fuzzier in nature than verification. Many definitions speak of assessing whether the computer program is "an accurate representation of the system under study,"[67] whether the simulation is a "good model of the target," [68] or a legitimate "representation of the actual…system under design or study." [69] According to these definitions, validation thus requires assessing the goodness of fit of the code to something external to the modeling process – the target domain. If we are interested in a 'fit' or isomorphism between a computational representation or model and a target domain, we can infer from common practice that we are interested in using the computational model to predict, to tell us what the future will look like. Hence this type of validation is more specifically referred to as predictive validation.

Predictive validation

Computational models in the physical sciences have historically been tested for their predictive capability. Prediction in a modeling regime can be defined as the "use of a computational model to foretell the state of a…system under conditions for which the computational model has not been validated."[70] This requires some means to compare the model to that target domain. In the physical sciences, engineering and, to some lesser degree, the life sciences, this is usually accomplished through experimentation.

William Oberkampf *et al* argue that validation is the process of determining the degree to which computational simulation results agree with experimental data.[71] (Oberkampf *et al*'s thorough

---

[65] U.S. Department Of Health and Human Services. Food and Drug Administration. Center for Devices and Radiological Health. Center for Biologics Evaluation and Research General Principles of Software Validation; Final Guidance for Industry and FDA Staff. Document issued on: January 11, 2002. P.5 http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm085371.pdf

[66] U.S. Department of Defense. Dictionary of Military Terms. http://www.dtic.mil/doctrine/dod_dictionary/

[67] Jack P.C. Kleijnen. 1995. Verification and validation of simulation models. European Journal of Operational Research. Vol.82. Pp.145-162

[68] Nigel Gilbert et al. 1999. Simulation for the Social Scientist. 1st edition. Philadelphia, PA: Open University Press.

[69] E. J. Williams. 1998.
Verification and validation in industrial simulation. Proceedings of the Summer Computer Simulation Conference. Pp. 57-62.

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

[71] William L. Oberkampf, Timothy G. Trucano, and Charles Hirsch. 2003. Verification, Validation, and Predictive Capability in Computational Engineering and Physics. Sandia Report. SAND 2003-3769. February 2003. U. S. Department of Energy, Sandia National Laboratories. Albuquerque, NM p.17

treatment of validation covers in depth many issues on which I will only touch in this discussion, although their discussion specifically targets the physical sciences.) In this paradigm, experimental data are used as a surrogate for the 'real world,' or as Oberkampf *et al* put it, "experimental data…[are] our best indication of reality."[72] Oberkampf *et al*'s statement that validation is about determining the degree of fit between the results of a simulation and the results of experiments raises the question of the degree of accuracy of that fit. Just how accurately does a model need to conform to the experimental results for the model to be 'valid'?

As Oberkampf and Matthew Barone note, "If the computational results "generally agree" with the experimental data, the computational model is commonly declared "validated."[73] Naomi Oreskes and Kenneth Belitz, speaking from the perspective of the earth sciences, remind us that "The literature of model validation is filled with terms like 'adequate', acceptable', 'satisfactory,' and reasonable', even in sophisticated mathematical treatments. These are obviously subjective terms…"[74] Leonard Konikow and John Bredehoeft point out that that "because the definition of 'good' [as in a 'good comparison' between experimental and simulation results] is subjective, under the common operational definitions of validation, one competent and reasonable scientist may declare a model as validated while another may use the same data to demonstrate that the model is invalid."[75]

While there are nascent efforts at the development of what are being called 'validation metrics (primarily in the physical sciences and engineering),[76] we will always be left with a process that yields high levels of uncertainty. As Karl Popper famously pointed out, "a positive decision can only temporarily support the theory, for subsequent negative decisions may always overthrow it."[77] So while the number of tests that exhibit some high degree of isomorphism between the model and the 'real world' or target system can increase, there is no way to determine with absolute certainty that the next test will also exhibit a corresponding degree of isomorphism. Continuing to test does provide other benefits, however. Thomas Naylor *et al* point out that "[i]f in a series of empirical tests of a model no negative results are found but the number of positive instances increases then our confidence in the model will grow step by step."[78] Therefore, we can *increase our confidence* in the likelihood that the model's output will conform to experiment results. But that is all.

Predictive validation and computational social models

---

[72] Ibid., p.13

[73] William. L. Oberkampf and Matthew F. Barone. 2005. Measures of Agreement Between Computation and Experiment: Validation Measures. Sandia Report. SAND 2005-4302. August 2005. U. S. Department of Energy, Sandia National Laboratories. Albuquerque, NM P.1

[74] Naomi Oreskes and Kenneth Belitz. 2001. Philosophical Issues in Model Assessment. In Model Validation: Perspectives in Hydrological Science. M.G.Anderson and P.D.Bates (eds.) John Wiley & Sons, Ltd. New York, NY. Pp.23-41. P.24

[75] Leonard F. Konikow and John D. Bredehoeft. 1992. Ground-water models cannot be validated. Advances in Water Resources. Vol.15, No.1. Validation of Geo-hydrological Models Part 1. Pp.75-83. P.78

[76]See, for example, Leonard E. Schwer. 2007. Validation metrics for response histories: perspectives and case studies. Engineering with Computers. Vol.23, No.1. Pp.295-309.

[77] Karl Popper. 2002 (1959) The Logic of Scientific Discovery. Routledge. P.10

[78]Thomas H. Naylor, J. M. Finger, James L. McKenney, William E. Schrank and Charles C. Holt. 1967. Verification of Computer Simulation Models. Management Science. Vol. 14, No. 2, Application Series. Pp. B92-B106

There are certain methodological issues related to predictive validation which makes it a very questionable process to apply to computational social models. These issues relate to the ethics of experimentation on human populations, the complex nature of the target domain, the nature of the data, and the time frames in question. I will treat each very briefly in turn.

Ethical limitations

The ethical limitations placed upon experimentation on human populations do not need much explication. The experiments in Tuskegee, TN in the 1950's where a 'control population' of African-American men was denied treatment for syphilis over a 40-year period although medication was available to the general population is a well-known example of an unethical experiment.[79] The 'Milgram experiment' in which subjects were willing to provide what they thought were fatal electrical shocks to subjects when so ordered by an authority figure[80] is another well-known example.

Obviously, however, we do experiment on human populations. These experiments are now conducted in accordance with one of several codes governing the conduct of experimentation on human populations. The two best known codes are the Nuremberg Code, promulgated in 1949 as a response to the actions of the Nazis on subjected populations during World War II, and the Belmont Report, issued in 1976 by the U.S. Department of Health, Education and Welfare and often referred to as the standard governing research funded by federal dollars.[81] Institutions managing research funds have established review boards that ensure that the research design and the conduct of research conform to these broad ethical guidelines.

The constraints imposed by complexity

Socio-cultural (human) systems are complex adaptive systems. These types of systems are composed of agents that are constantly learning and adapting to their environment. In the course of this evolution, not only do the agents change but the structure of the system (the organization of the agents) also changes. One important consequence of this is that the system as an aggregate will exhibit behavior that is the outcome of many decisions made at the local level, a phenomenon that has come to be known as 'emergence.' As John Holland put it, "[b]ecause the individual parts of a complex adaptive system are continually revising their ("conditioned") rules for interaction, each part is embedded in perpetually novel surroundings (the changing behavior of the other parts).[82] This also contributes to the nonlinearity that Sallach (this volume) describes as an important aspect of the complexity of social systems.

The complex systems nature of human systems has two consequences for the development of computational models. First, if a part of the system is isolated for experimental purposes for a validation exercise, the integrity of the system as a whole has been compromised. The *parts cannot be understood outside of the context of the system as a whole.* Sallach (this volume) treats this in terms of what he calls endogeneity, noting that "whether an attribute is exogenous or endogenous is usually regarded as a characteristic of the *model* rather than the social process that is being represented." Secondly, emergence – "a phenomenon whereby well-formulated aggregate behavior arises from

[79] http://www.cdc.gov/tuskegee/timeline.htm

[80] Stanley Milgram. 1974. Obedience to Authority; An Experimental View. Harpercollins. New York, NY

[81] Both codes can be found at http://ohsr.od.nih.gov/guidelines/index.html

[82] See John. H. Holland. 1992. Complex Adaptive Systems. Daedalus, Vol. 121, No. 1. Pp. 17-30. P.20

localized, individual behavior"[83] – means that producing a model that reproduces the behavior of the target system does not mean that the model accurately represents the structure or underlying processes of that system. It thus is no assurance that the model is predictive, i.e. that it will reproduce the target system's behavior under different conditions.[84] As Holland says, "History and context play a critical role... It is the process of becoming, rather than the never-reached end points, that we must study if we are to gain insight."[85]

Data issues

Aaron Perls and I have presented an extensive argument elsewhere about some of the problems with sociocultural data related to computational modeling.[86] Because all entities and relationships must be able to be treated computationally, they must be observable and quantifiable. Not all human phenomena fit this description. Belief, for example, is a very strong sociocultural factor and yet cannot be either seen or counted. These phenomena are usually treated in the computational domain through surrogates – behavior substituting for belief or motivation, for example. This substitution is often not (or poorly) documented, and its impact on model utility rarely recognized.

There are additional issues with sociocultural data, many revolving around uncertainty. There is epistemological uncertainty (have we collected the right data), as well as uncertainty related to collection and measurement (have we collected the data right) and its conversion to computationally manipulable forms. There is a huge literature related to uncertainty which I will not even begin to treat here. However, the recognition of its role in computational social modeling has heretofore been rather underplayed. Suffice it to say that a National Research Council report on dynamic social network analysis identified uncertainty as one of the key under-researched areas in quantitative and computational social science[87]

As a final note in this section, we refer to Robert Albro's work in this volume in which he challenges the very notion of 'data' itself. He argues that the operational and use requirements the military places upon the modeling of sociocultural phenomena generate, in effect, a new type of sociocultural data. Although the term may be the same as that used by anthropologists in their ethnographic work, the 'data' upon which concepts such as 'human terrain' are based are epistemologically different. As Albro puts it, "The relevance of cultural data for the user community [in this context] appears to be more important than its significance for ostensible cultural subjects themselves." He thus problematizes the very notion of data itself.[88]

---

[83] John H. Miller and Scott E. Page. 2007. Complex Adaptive Systems: An introduction to computational models of social life. Princeton University Press. Princeton, NJ. P.46

[84] Oreskes and Belitz, op.cit. P.32

[85] Holland, op.cit.

[86] Jessica Glicken Turnley and Aaron S. Perls. 2008. What is a computational social model anyway?: A Discussion of Definitions, a Consideration of Challenges, and an Explication of Process. U.S. Department of Defense. Defense Threat Reduction Agency, Advanced Systems and Concepts Office. Report No. ASCO 2008-013. http://www.au.af.mil/au/awc/awcgate/dtra/what_is_social_model.pdf

[87] R. Breiger, K. Carley, and P. Pattison. 2003. Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers. Washington, DC: National Academies Press.

[88] I would go a bit further and suggest that there may not be a *correct* definition of data here, but rather an *appropriate* one. The challenge is to ensure that all participants in the process are using the *same* definition.

Dealing with time

There are several ways in which time makes validation of computational social models difficult. Perhaps most important is that of scale. Sallach (this volume) speaks of the "unity of time horizon" that is absent in the nonlinear systems of which social systems are an example. Changes in human systems often take place over generations. Large-scale changes may not be evident for centuries. Therefore, in order to see if a model is 'predictive,' i.e. if its output is isomorphic with what 'really happens,' we would need to live long lives ourselves.

Ecosystem and climate change models face the same challenge. The inability to assess the predictive validity of climate change models, for example, is one of the many issues that have led to the long debate on the credibility of the threat. Edward Rastetter proposes an approach (not a solution) to this problem. "Confidence in these models has to be built through the accumulation of fairly weak corroborating evidence rather than through a few crucial and unambiguous tests. The criteria employed to judge the value of these models are thus likely to differ greatly from those used to judge finer scale models, which are more amenable to the scientific tradition of hypothesis formulation and testing." [89]

Another time-related problem that ecosystem models and models of socio-cultural systems have in common is that of multiple timeframes impacting the same problem.[90] In the ecosystem arena, for example, any given ecosystem is experiencing phenomena that change according to geologic and climactic time scales (eons, with occasional rapid convulsions), ecological scales (decades), weather-related (as distinct from climatic) scales (daily), and so on. The same is true for human communities. Some phenomena change quickly (fashion, certain types of behaviors, and the like). Others take generations to change, while still others may change even more slowly. Accounting for the interaction of these different rhythms is a significant challenge for a computational social model.

Other types of validation

These issues suggest that computational social models may, in fact, be something other than computational physics models using a different set of data. Thinking of them as redirected physics models may force us to ask of them that which they cannot produce. So, if not validation as we think of it for computational models of physical or biological phenomena….then what?

The question should actually be, 'if not *predictive validation*…then what? We may not be ready to completely dispense with validation as predictive validation is only one type of validation exercise that could be performed on a computational model. There are others.

Paul Davis, in his discussion of simulations for military uses, identifies three different types of validity: predictive validity, descriptive validity, and structural validity. He recognizes but does not privilege predictive validity, which he defines as existing when "a model…can predict desired features of system behavior, at least for particular domains of the initial conditions and durations of

[89] Edward B. Rastetter. 1996. Validating Models of Ecosystem Response to Global Change: How can we best assess models o f long-term global change?. BioScience Vol. 46 No. 3. Pp.190-198. P.190

[90] Simon A. Levin. 1992. The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture Ecology, Vol. 73, No. 6. Pp. 1943-1967.

time, to within some known level of accuracy and precision."[91] Note that his notion of predictive validity is highly constrained. Any given model may have such validity but only under certain specified constraints.

Davis also introduces a concept of descriptive validity. This is present when "the model is able to *explain* phenomena or organize information meaningfully in one way or another"[92] (emphasis in the original). And finally, he also talks about structural validity, which "means that the model has the appropriate entities (objects), attributes (variables) and processes so that it corresponds in that sense to the real world (verisimilitude)."[93]

Lance Champagne, in his exploration of verification and validation for certain combat simulations, introduces another set of validity techniques. He speaks of empirical validity, face validity, and theoretical validity. Empirical validity is what I have called predictive validity: "the aim of empirical validity techniques is to provide an indication as to the accuracy of the model with respect to the observed behavior of the system under study."[94] Face validity is established when the model "appears reasonable" to those with subject-matter knowledge.[95] (Face validity has been called 'social validity' in the literature on participatory or companion modeling.[96] I will discuss this literature in more depth later.) And theoretical validity is used to "establish the extent to which a model conforms to scientific theory."[97]

I would take issue with his last point. In our extensive treatment of the nature of models elsewhere,[98] we argued that this process – the comparison of the computational model to theory – is actually a form of verification. We have treated elsewhere what it means to 're-present' a target system through a computational model.[99] A model is not the whole system, for then it would not be the model – it would be the system. Briefly, we argued that a model is an analogy of the target system. The modeling team identifies what it considers to be salient elements and relationships in the target environment and re-produces them computationally. Clearly, it is the definition of salience that is in question here, for it is this definition that forms the logic of choice. (Possibly our 'salience' is what Davis meant by 'appropriate' in his definition of structural validity, where the model has the "appropriate entities (objects), attributes (variables) and processes"[100]). We argued that such salience is determined by the questioner (a particular social role in the model building

---

[91] Paul K. Davis. 1992. Generalizing Concepts and Methods of Verification, Validation, and Accreditation (VV&A) for Military Simulations. Prepared for the Under Secretary of Defense for Acquisition. RAND R-4249-ACQ. RAND Corporation. Santa Monica, CA. P.8

[92] Ibid. P.7

[93] Ibid.

[94] Lance E. Champagne, MAJ, USAF. 2003. Development Approaches Coupled with Verification and Validation Methodologies for Agent-Based Mission-Level Analytical Combat Simulations. Dissertation. Graduate School of Engineering and Management. Air Force Institute of Technology. Air University. United States Air Force. P.60

[95] Ibid., P.58

[96] Castella, Jean-Christophe, Tran Ngoc Trung, and Stanislas Boissau. 2005. Participatory Simulation of Land-Use Changes in the Northern Mountains of Vietnam: the Combined Use of an Agent-Based Model, a Role-Playing Game, and a Geographic Information System. Ecology and Society Vol.10. No.1 Pp 27ff.

[97] Champagne, op.cit. P.62

[98] Turnley and Perls, op.cit.

[99] Ibid.

[100] Davis, op.cit. p.7

process), usually in conjunction with the model user (another role), and is expressed in the model through the theory that underpins it. It is particularly important to make visible the use of theory as the logic of selection in the computational social science world, for each set of phenomena in the target domain (the 'real world' of human interaction) has a myriad of explanatory theories.

Each theory drives the modeling team to select a different set of objects and relationships as relevant. Once those are identified, the programmer writes code to instantiate them in the computational environment. Comparing the code to the theory, in this case, is actually determining if the code does what the programmer intended it to do – verifying the code, in other words, not validating it. Validation is performed on the conceptual model or theoretical construct. Is an economic explanation the 'correct' or appropriate one? Or is the phenomenon better explained by structural theory? Or perhaps the issue is one of affective engagement – invoking yet another disciplinary body of knowledge and making salient (or appropriate) yet another set of objects and associated relationships? (I note here for the record that there are some interesting ways in which the interplay between the code and theory can extend the theory. The code actually is more than the theory expressed in a different language. However, a full discussion of this would take us off topic – and at the level of resolution of this paper, my earlier statements still stand.)

At best, we might say that narrative social theory developed through observation and/or limited experimentation is validated for its predictive capability through case studies, ethnographies and the like. Computational models structured and shaped by narrative theory use the validity of that underlying theory as a proxy for their own.

So we now have the following:

- A validation method that tests the 'fit' between the model and the target domain (the 'real world') – *predictive validation.*
- A validation method that assess whether the model organizes information about a phenomenon in a meaningful way, allow us to learn something new about that phenomenon – *descriptive validation.*
- A validation method that checks how well the computational model represents a theory – *structural or theoretical validation*, a method I claim is actually verification not validation.
- A validation method that checks whether or not the model 'looks reasonable' to people who know something about the target domain – *face or social validation.*

There may be other types of validation as well. My point is that predictive validation is an important method for assessing the goodness of a model *if, and only if,* the purpose of the model is to predict. As Davis pointed out, "…evaluation of models should vary with type. It is silly to denigrate a good descriptive model that is structurally valid, merely because it is not a prediction machine…This is nontrivial, because many critics of military modeling are guilty of precisely this error."[101] And, as we saw earlier, there may be certain aspects of the socio-cultural domain that make it very difficult if not impossible to create a computational social model with a high level of predictive capability.

Models and their Uses

Just because computational social models cannot predict does not mean they are useless. This, then, begs the question of the purpose or utility of such models. Laura McNamara argues that "the main

[101] Davis, op.cit. P.10

benefit of V&V is not (perhaps counter-intuitively) increased focus on the model but the contextual issue of how the model will be used and, therefore, how the organization and its members identify what decisions they are responsible for making and how they negotiate acceptable levels of risk. This is because verification and validation emphasize whether or not a software application is credible *for an intended area of use*" (emphasis in original).[102]

Not surprisingly, we find that organizations which utilize software in furtherance of their missions add a dimension of 'usefulness' to their definitions of validation. The U.S. Food and Drug Administration, when speaking of requirements for software used in medical applications, defines validation as the "confirmation by examination and provision of objective evidence that software specifications conform to user needs and intended uses, and that the particular requirements implemented through software can be consistently fulfilled."[103] The U.S. Department of Defense defines validation as "the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model."[104]

Models answer some question or, to put it another way, fill some need. That question may be theoretical (do the posited cause and effect relationships hold under certain specified conditions, often phrased as an if-then question?), it may be descriptive (what does a certain phenomenon look like?) or it may be applied, designed to contribution to the solution of some 'real world' problem, as in the case of military, corporate and policy models. Hodges, in his rather tongue-in-cheek description of uses for a 'bad model' (i.e. one that does not predict well), identified uses ranging from a rather mundane (but nonetheless important) bookkeeping function, to a more substantive use such as an aid to thinking and a stimulus to intellectual exploration.[105]

Assessing the Utility of a Model

So how do we track the 'utility' or measure the usefulness of a computational model? We certainly could look at the number of times it is used (as with a webpage counter) and by whom. More meaningfully, we could assess the contribution the model makes to its stated use domain. In order to do this, we need to a) clearly understand how the modeling team characterized the use domain (i.e. the model's stated purpose or intended use), and b) understand how a tool such as a model can achieve that purpose or be used in its intended manner.

One way to think of the use of computational models is to focus on model output, the product of a model run. This output could be used in two ways. It could be seen as the answer to the question, the solution to a problem, i.e. the end state of the knowledge-seeking process. It also could serve as input to a decision made by a human. I will briefly treat the former case, but spend more time on the latter as I will argue that it is as input to human decisions that computational social models can make their most valuable contributions.

---

[102] Laura McNamara. 2010. Why Models Don't Forecast. Draft paper prepared for the workshop Unifying Social Frameworks. National Research Council, Washington, DC. August 16-17, 2010. http://www7.nationalacademies.org/bbcss/Why%20Models%20Dont%20Forecast-McNamara.pdf P.10

[103] U.S. Department of Health and Human Services, op.cit. P.5

[104] U.S. Department of Defense. Dictionary of Military Terms.

[105] James S.Hodges. 1991. Six (or so) things you can do with a bad model. Operations Research. Vol.39. No.3. Pp.355-365. Pp. 355 - 6

Computational problem-solving systems are often termed 'expert systems.' There is a voluminous literature on expert systems which I will not treat here, other than to point out that expert systems only work well in highly constrained environments. "Expert systems are typically autonomous problem-solving systems used in situations where there is a well-defined problem and expertise needs to be applied to find the appropriate solution"[106] These environments are generally those in which there is a possible answer, a 'right' answer. Furthermore, while an expert system (computational or otherwise) can describe the steps necessary to get to a solution, it generally provides little explanation of the solution logic to the user.[107]

In complex systems such as socio-cultural systems, problems are rarely well-defined nor do they deal with highly constrained environments. In part this is because we are still ignorant of many of the variables that affect a given problem space. It also is because often simply the act of framing the question affects the target situation. Think, for example, of an exercise to evaluate a site for potential Superfund status. Part of the evaluation is a socio-cultural assessment of the affected community. However, simply the announcement that the evaluation will take place (as distinguished from actually performing the evaluation) could affect property values, cause some families to move out of the neighborhood, affect the availability of commercial capital, and the like. In this type of situation, the full set of parameters defining the problem may be unknown, and for those that are known, their value may be constantly changing. These types of problems have recently been discussed in some circles as 'wicked problems.'[108] From a more useful analytic perspective, they also may be treated as a function of the complex nature of the system. As Holland pointed out, "the aggregate behavior of the system is usually far from optimal, if indeed optimality can even be defined for the system as a whole. For this reason, standard theories in physics, economics, and elsewhere, are of little help because they concentrate on optimal end-points, whereas complex adaptive systems "never get there." They continue to evolve, and they steadily exhibit new forms of emergent behavior."[109]

So suppose we move away from the conception of a computational model as 'making a decision' or providing an answer, to one in which the model acts in a decision support role. Computational models operating in a decision support role can be characterized as advisors. "Advisory systems do not make decisions but rather help guide the decision maker in the decision-making process, while leaving the final decision-making authority up to the human user."[110] As McNamara pointed out, "models don't forecast because *people* forecast" (emphasis in original).[111]

This type of system, in which a human makes a decision using a process which incorporates input from many different sources, is often called a 'judge-advisor' system. "….most important decisions are made in a social context in which the person responsible for the decision, the Judge, seeks or receives some from (*sic*) of input from one or more persons in the role of Advisor…the premise is

---

[106] Brandon A. Beemer and Dawn G. Gregg. 2008. Advisory Systems to Support Decision-making. In Handbook on Decision Support Systems 1. International Handbook on Information Systems, Springer. Pp.361-377. P.361

[107] G Luger. 2005. Artificial Intelligence: Structures and Strategies for Complex Problem Solving. Addison Wesley, cited in Beemer and Gregg, op.cit. P. 365

[108] W. Horst, J.Rittel and Melvin M.Webber. 1973. Dilemmas in a General Theory of Planning. Policy Sciences 4, 155-169

[109] Holland, op.cit.

[110] Beemer and Gregg, op.cit. p.361

[111] McNamara, op.cit. P.4

that the multiple persons involved in making a decision do so in very different roles."[112] In this framing, while the ultimate decision is made by a single individual (the 'judge'), that decision is partially a function of the advice received. The advisors thus are present in that final decision and must be considered in order to understand how the final decision is made.

I am positing that a computational social model most effectively serves in an advisory capacity, providing two types of input to a judge's decision. One type involves the provision of clarity or information about the structure or logic of the complex system about which the decision is to be made. The other type is the provision of possible futures for consideration.

One strong area of value of computational models over narrative or qualitative models is that all entities and relationships in a computational model must be made explicit, all parameters stated clearly, and all dynamics unambiguously described for the code to be written. This forces rigor upon the narrative world of social theory (although I emphasize that the multivocality and indeterminacy of narrative does have an important role to play for certain purposes). It also contributes to understanding by making visible what may have been unseen, allowing (forcing) us to engage with it. As Robinson Crusoe tells us upon seeing the footprint of an unknown person on the sands of his desert island, "Not that I did not believe that savages had frequented the island even all the while…but I had never known it, and was incapable of any apprehensions about it; my satisfaction was perfect, although my danger was the same, and I was as happy in not knowing my danger as if I had never really been exposed to it."[113] Recall Davis' explanation of descriptive validity: "the model is able to explain phenomena or organize information meaningfully in one way or another."[114]

Different types of computational social models fill this explanatory role differently. Systems dynamics models, for example, help us "understand …how all the objects in a system interact with one another"[115] These types of models are directly focused on providing a better understanding of overall system behavior by identifying objects and the dynamics of their relationships. The focus is on the dynamics of the system. Social networks also aim to show us how structure can influence group behavior by allowing us to add or remove nodes (actors) or links (connections) and providing snapshots of the resulting different structural arrangements. The focus here is on the relationships or connections between nodes. Agent-based models illustrate how small perturbations in a system at a very low level can have large, system-wide effects. In this modeling approach, the focus is on describing the actors (including the attributes which allow for differentiation among them) and the dynamics ('rules') of their interactions.

All three modeling approaches require the identification of entities (objects, actors) in the system and connections among them. In the case of systems dynamics and agent-based models, they also must describe dynamics or process, the ways the entities interact over time. In the case of social networks, there are no dynamics incorporated into the model itself but the use of repeated 'snapshots' of different configurations can provide information on how the structure can change over time (although not why it changed). In all three approaches, the process of identifying and

---

[112] Janet A. Sniezek . 1999. Judge Advisor Systems Theory and Research and Applications to Collaborative Systems and Technology. Proceedings of the 32nd Hawaii International Conference on System Sciences – 1999. IEEE. p.1

[113] Daniel Defoe. 2008 (1719). Robinson Crusoe. Serentity Publishers. Pp.152-153

[114] Davis, op.cit. P.7

[115] http://sysdyn.clexchange.org/sd-intro/home.html

describing these entities, the relationships among them, and the processes that can affect system functioning can themselves be a learning process.

Participatory modeling (also known as companion modeling) is a mechanism in which stakeholders (who are often also decision-makers) engage in a role-playing game or some other mechanism through which modelers elicit their tacit knowledge about the rules for interaction in a pre-defined situation. These rules are then encoded in an agent-based model which is re-presented to the players (stakeholders, decision makers) for face or social validation.[116] Facilitators of this type of participatory process claim that it is highly educational for the participants, providing them with knowledge and insight into other players' agendas that they otherwise would not have.[117] This benefit was particularly emphasized by those who took on roles other than their own during the course of the game.[118] The game also produces knowledge about how players move through a particular complex system.[119] As Thomas Karas pointed out in his discussion of a workshop focusing on the relationships between modelers and policymakers, "[p]articipatory modeling can benefit those involved in building the model even when the completed model is not used as a communications framework. The very act of constructing a model requires learning, in a structured way, how the modeled system works. Thus the mental model that the participant leaves with may be more sophisticated and more reflective of the best knowledge on the subject—the analyst or policymaker becomes a more proficient expert himself."[120] In this way, "modeling has the potential to enhance the [policy] process by…deepening policymakers' comprehension of the underlying problems and issues, clarifying decision-makers' assumptions and values helping to build understandable narratives ("stories") in support of policy proposals, informing dialogue among stakeholders and policymakers, or providing a framework for negotiation and consensus building.[121] The model allows us to 'see' what has been hidden, to see Robinson Crusoe's 'savages' as revealed by their footprints on the sand.

Clearly in this case of assessment of model goodness, an assessment of the goodness of the computational model, would involve not a direct evaluation of the model, but an assessment of its impact on the model user. For example, pre- and post-test instruments could be administered to gauge the knowledge or understanding delta exhibited by the user as a result of his engagement with the model.[122] A 'good' model will generate a large delta.

Computational social models can be used to generate alternate scenarios or possible futures for decision makers which can also serve as a learning mechanism. With a computational model, the

---

[116] Marco A. Janssen and Elinor Ostrom. 2006. Empirically Based, Agent-based models. Ecology and Society 11(2): 37. Accessed online at http://www.ecologyandsociety.org/vol11/iss2/art37/

[117] Dionnet, M., M Kuper, A Hammani, P Garin. 2008. Combining role-playing games and policy simulation exercises: an experience with Moroccan smallholder farmers. Simulation Gaming OnlineFirst. Published February 19, 2008

[118] Gurung Tayan Raj, Francois Bousquet and Guy Trébuil. 2006. Companion Modeling, Conflict Resolution, and Institution Building. Ecology and Society Vol. 11, No.2. Pp36ff Accessed online at http://www.ecologyandsociety.org/issues/search.php

[119] Gurung *et al*, op.cit.

[120] Thomas Karas. 2004. Modelers and Policymakers: Improving the Relationships. Sandia report. SAND 2004-2888. U.S. Department of Energy. Sandia National Laboratories. Albuquerque, NM. P.12

[121] Ibid. Pp.6-7

[122] This, of course, begs the question of how we measure 'learning' – a road down which we will not go for this discussion.

decision maker can alter input variables and assess the corresponding difference in outcome, with the computer able to accommodate the permutation of these differences through more variables and relationships than the mind can handle. In this case, the decision maker *is* interested in model output – but not in model output as 'fact' or as a future that *will* happen, but in model output as possibility, as possible manifestations of the interaction of underlying structures and their dynamics. The interest should be not in the actual future but in the possibility space illustrated by the outputs of multiple runs which themselves are the function of changes in input variables.

As an interesting parallel, we might note the way in which Aristotle differentiated between history and poetry [art/literature]: "The real difference is this, that one [history] tells what happened and the other [poetry] what might happen. For this reason poetry is something more scientific and serious than history, because poetry tends to give general truths while history gives particular facts.[123] Mark William Roche, writing on the importance of literature in the 21st century, argued that literature works with something akin to Max Weber's ideal type.[124] As Roche says, "Goethe's Faust captures more of what life means than may a 'real' scholar or lover. In a sense he [Faust] is the original; we may appear to be like or unlike him, not he like or unlike us."[125] Literature thus may be true, although not accurate.

The presentation of a multitude of possible scenarios to a decision maker (the demarcation of a possibility space) begs the question of what the decision maker does with that possibility space. Answering this question would take us back to a consideration of what types of scenarios are most useful, and how they can be best communicated to the decision maker. This would take us into the literature on decision-making and scenario planning, a step too far for this discussion. For our purposes here, it is sufficient to note that this alternative approach to the assessment of goodness of the model once again moves away from the model's fit to the 'real world,' with the exception perhaps, an assessment of face or social validity to ensure that the computational model passes a 'laugh test.' As McNamara points out, "These discussions [of V&V] force clarification about the decisions, tradeoffs, and risks across stakeholder communities, and what is required for a model to be considered credible and appropriate in relation to a decision."[126] We might recall our parallel from literature here.

We still read and learn from Shakespeare (and Goethe, for that matter), although his world is no longer with us. We learn from Crusoe's encounter with footsteps in the sand, though we probably will never be stranded on a desert island. We abstract from the literature the archetypical elements of people and social interactions, and watch how those play out in a range of social settings. We then apply those lessons to our own lives, making (we hope) better informed social engagements.

Conclusion

Our argument is based on the claim that computational social models are inherently unable to be used for predictive purposes in the same way that models of physical and many biological phenomena can. Ethical issues, the complex nature of the socio-cultural domain, the nature of the

---

[123] Aristotle. Poetics. 1451b. accessed online at http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.01.0056%3Asection%3D1451b

[124] Mark William Roche. 2004. Why Literature Matters in the 21st Century. Yale University Press. New Haven. P.19

[125] Ibid.

[126] McNamara, op.cit.

data, and the inherent time scales of the phenomena under consideration severely constrain the types, breadth and depth of experiments we can conduct to ascertain predictive capability. At best, computational social models can rely upon the (proxy) validity of the narrative theory upon which they are based.

That said, there are other means of ascertaining goodness. All computational models are built with a use in mind. If we broaden our thinking and consider ways other than prediction to use computational social models, we will begin to realize alternative means of assessing their goodness. The question of degree of isomorphism to the 'real world' (or target domain) may become moot. Just as we continue to read Goethe although we never expect to actually meet Faust, so might we build and use computational social models although we never expect to actually live in the future any one of them 'predicts.' Although not accurate, computational social models can still be useful – and, even better, true.

APPENDIX D: PARTICIPANT BIOGRAPHICAL STATEMENTS

Participant Biographical Statements

Robert Albro received his PhD in sociocultural anthropology from the University of Chicago in 1999. Since 1991 Dr. Albro has maintained long-term ethnographic research, and published widely, on popular and indigenous politics in Bolivia, with a particular focus on the changing terms of citizenship and indigenous identity in this country. His current research is concerned with global cultural policy making, as it meaningfully shapes the ongoing terms of globalization, in particular, applications of the culture concept in contexts of security. Over the years, Dr. Albro's research and writing have been supported by the National Science Foundation, Mellon Foundation, Rockefeller Foundation, and the American Council for Learned Societies, among others. Dr. Albro has also been a Fulbright scholar, and has held fellowships at the Carnegie Council for Ethics in International Affairs, the Kluge Center of the Library of Congress, and the Smithsonian Institution. Dr. Albro has held several leadership positions in the American Anthropological Association, including most recently as Chair of the Committee for Human Rights and Chair of the Ad Hoc Commission on Anthropology's Engagement with the Security and Intelligence Communities. He was given the AAA's President's Award in 2009 for outstanding contributions to the Association. Dr. Albro is currently a member of the National Research Council's Committee on Unifying Social Frameworks and teaches in the School of International Service at American University.

Mark Bevir is a Professor in the Department of Political Science at the University of California, Berkeley. His research interests include political theory and public policy. He is the author or co-author of *The Logic of the History of Ideas* (1999), *Interpreting British Governance* (2003), *New Labour: A Critique (2005), Governance Stories* (2006), *Key Concepts in Governance* (2009), *The State as Cultural Practice* (2010), and *Democratic Governance* (2010). He is the editor of two reference works – *The Encyclopedia of Governance* (2007) and *The Encyclopedia of Political Theory* (2010) – and the editor or co-editor of *Critiques of Capital in Modern Britain and America* (2002), *Markets in Historical Contexts: Ideas and Politics in the Modern World* (2004), *Modern Political Science: Anglo-American Exchanges since 1880 (2007), Public Governance* (2007), *Histories of Postmodernism* (2007), *Governance, Consumers, and Citizens: Agency and Resistance in Contemporary Politics* (2007), *Interpretive Political Science* (2010), and *The Sage Handbook of Governance* (2010). Mark has also published over 150 articles in books and journals.

Gerald Epstein joined the American Association for the Advancements of Science (AAAS) Center for Science, Technology, and Security Policy (CSTSP) as Director in October 2009. Prior to joining CSTSP, he was Senior Fellow for Science and Security in the Homeland Security Program at the Center for Strategic and International Studies, where he worked on reducing and countering biological weapons threats and improving relations between the scientific research and national security communities. Dr. Epstein has also held positions with the Institute for Defense Analyses; the White House Office of Science and Technology Policy (OSTP), where he served jointly as Assistant OSTP Director for National Security and Senior Director for Science and Technology on the National Security Council staff; and the Congressional Office of Technology Assessment. He also directed a project at Harvard University's Kennedy School of Government on the relationship between civil and military technologies and is a co-author of *Beyond Spinoff: Military and Commercial Technologies in a Changing World* (Boston, MA: Harvard Business School Press, 1992). Dr. Epstein has served as visiting lecturer in public and international affairs at Princeton University's Woodrow Wilson School, and he taught "Science, Technology, and Homeland Security" as an adjunct professor with Georgetown University's Security Studies Program. Dr. Epstein is a Fellow of the American Physical Society and the American Association for the Advancement of Science, and he is

a member of the Biological Threats Panel of the National Academy of Sciences' Committee on International Security and Arms Control, the Biological Sciences Experts Group for the Office of the Director of National Intelligence, and the editorial board for the journal *Biosecurity and Bioterrorism*. He had served on the National Academies' Committee on Science, Security, and Prosperity. He received S.B. degrees in physics and in electrical engineering from MIT and a PhD in physics from the University of California at Berkeley.

Charles Gieseler is a software engineer who is interested in human computer interaction and computational modeling. He completed a Masters in Computer Science from Iowa State University in 2005 with a focus in machine learning for agent-based computational economics. He is currently a software engineer working with the Cognitive Modeling department at Sandia National Laboratories where he is developing user interaction and simulation technologies for cognitive modeling applications.

François M. Hemez has been Technical Staff Member at Los Alamos National Laboratory since 1997. He was a member of the Weapon Response group (ESA-WR) for seven years; served as ESA-WR Validation Methods team leader for one year; and is currently with XCP-Division. He has managed the verification project of the Advanced Scientific Computing program for two years and currently manages the predictive capability assessment project, while contributing to the development and application of Verification and Validation (V&V), uncertainty quantification and decision-making for engineering, nuclear energy and weapon physics projects. François co-founded and chaired the Society for Experimental Mechanics (SEM) technical division on model validation and uncertainty quantification (2005-2009); served on the SEM executive board (2007-2009); and has been serving on the advisory board of the SEM International Modal Analysis Conference since 2006. He developed and taught the first-ever graduate course offered in a U.S. University (University of California San Diego, spring 2006) in the discipline of V&V. François received the Junior Research Award of the European Association of Structural Dynamics in 2005; received two U.S. Department of Energy Defense Program Awards of Excellence for applying V&V to programmatic work at LANL in 2006; and was the recipient of the SEM D.J. DeMichele award in 2010. François has authored over 300+ publications and reports (including 23 peer-reviewed papers) since 1994.

Phillip Huxtable received his PhD in Political Science in 1997 from the University of Kansas, with a focus on African Politics and quantitative analysis. Since joining the U.S. Department of Defense in 1999, he has filled a variety of roles, leading teams focused on incorporating the theories and methods of the social, engineering, and physical sciences into analysis in support of military operations across the spectrum, including peace-time operations through conflict and into post-conflict reconstruction. In 2005, he was formally recognized for his work adapting the methods of Social Network Analysis into techniques appropriate for the analysis of national security threats. Dr. Huxtable currently leads a team of senior scientists and engineers charged with assessing emerging analytic technologies for their potential relevance to future national security issues.

Jeffrey C. Johnson is a Senior Scientist at the Institute for Coastal Science and Policy, and University Distinguished Research Professor in the Department of Sociology with adjunct appointments in Biology, Anthropology, and Biostatistics at East Carolina University. He received his PhD in Social Science from the University of California Irvine in 1981. He is also the Social Science Program Manager for the U.S. Army Research Office where he is developing a basic scientific research program in the social sciences. He has been actively involved in long-term

research, funded by the National Science Foundation, on human cognition and social networks. This includes a 10 year study of group dynamics in Antarctic research stations in which he developed and tested network models of the evolution of network structure and its relation to group outcomes (e.g., morale and performance) involving the American South Pole station, Russia's Vostok station, China's Great Wall station and Poland's Arktowski station. In addition, he has worked on developing cultural ecological models of Inupiaq cultural knowledge of ecosystem function and dynamics in the arctic and cultural models of fairness and justice in China. This reflects a general interest in development of more formal models of culture and means for gaining a better statistical understanding of factors underlying intracultural variation. Some of his more recent work has involved formal network models of food webs and food web dynamics. He is director of the Summer Institute for Research Design in Cultural Anthropology (funded by the National Science Foundation), co-editor of the journal *Human Organization* and associate editor of the *Journal of Social Structure*.

Laura A. McNamara is a Principal Member of Technical Staff in the Exploratory Simulation Technologies Organization at Sandia National Laboratories. Trained in cultural anthropology, McNamara conducts field studies in national security environments to assess barriers and opportunities for new technology development and adoption. She has worked with nuclear weapon experts, intelligence analysts, and cybersecurity experts, focusing on issues of expert knowledge elicitation and representation, verification and validation in computational social science, uncertainty quantification, user centered design strategies, innovation adoption, and software evaluation. Dr. McNamara received her PhD in Cultural Anthropology from the University of New Mexico in 2001, with a dissertation on the problem of knowledge loss in the post-Cold War nuclear weapons programs at Los Alamos National Laboratory (LANL). After completing her doctorate, she spent three years as a staff member in LANL's Statistical Sciences Group before moving to Sandia in 2004. Over the past decade, Dr. McNamara has worked with the Missile Defense Agency, the Defense Intelligence Agency, and the nuclear weapons programs at Sandia and Los Alamos National Laboratories to identify both the limitations of, and leverage points for, the effective use of modeling and simulation technologies in interdisciplinary research and development projects. As the team lead for Sandia's Networks Grand Challenge, she is currently working on evaluation strategies to determine how novel information visualization techniques impact knowledge production in intelligence organizations. Dr. McNamara is a Fellow of the Society for Applied Anthropology and spent two years on the American Anthropological Association's Commission on the Engagement of Anthropology with the U.S. Security and Intelligence Communities. McNamara was recently appointed to the National Research Council's Committee on Improving the Decision-making Abilities of Small Unit Leaders (2010-2011). In addition, she serves on Sandia's Human Studies Board. She is co-editor (with Robert Rubinstein) of the forthcoming book, *Dangerous Liaisons: Anthropologists and the National Security State*, to be published by SAR Press in 2011.

Jennifer Perry is an analyst in the Defense Threat Reduction Agency's (DTRA) Advanced Systems and Concepts Office (ASCO). She leads ASCO's research activities on African security challenges as a principal investigator and/or project manager. She also provides technical guidance on current and prospective projects in the social sciences and currently engages with the academic, governmental, and contractor-based scientific communities to conduct research to identify and address the challenges of using computational social modeling for national security decision support. Jennifer also serves as the office expert on small arms and light weapons proliferation threats and issues. In 2005, she was seconded to DTRA's On-Site Inspections Directorate to support small

arms and light weapons threat reduction efforts and to the U.S. Department of State's Bureau of African Affairs to support public diplomacy efforts in the areas of counter-terrorism, trade, and development. Prior to joining DTRA, Jennifer served as a research intern at the Chemical and Biological Arms Control Institute and as a public diplomacy intern in the U.S. Department of State's Bureau of European and Eurasian Affairs. She also supported development activities in the international disability rights arena by monitoring, researching, and analyzing initiatives in Spanish and Portuguese-speaking countries at the Inter-American Institute on Disability. Additionally, she served as a conference reporter for the United Nations plenary and working group sessions toward a Convention on the Dignity and Promotion and Protection of Rights of Persons with Disabilities for Landmine Survivors' Network. Jennifer earned her B.A. in Sociology and Spanish (Linguistics) from Houghton College in Houghton, NY and her M.A. in International Affairs from the School of International Service at American University in Washington, DC, where she focused on international security and ethnic conflict.

Lucy Resnyansky is Research Scientist with the Defense Science and Technology Organization, Australia. She has a Bachelor (Honors) degree in Linguistics (1985) and a PhD in Social Philosophy (1994) from Novosibirsk State University (Russia); and a PhD in Education (2005) from the University of South Australia. She is an Affiliate Researcher with the Centre for Research in Education (CREd), University of South Australia. Her research interests are in the areas of social modelling, interdisciplinary research; sociocultural implications of technology; the Internet-mediated social interaction; activity theory; and social semiotics.

David Sallach received his PhD from the University of Nebraska at Lincoln. He taught sociology at Indiana University, Bloomington, Washington University, Saint Louis and the University of Missouri, Columbia. He taught computer science at the University of Nebraska, Kearney and the University of Arkansas Fayetteville. He has also provided information technology consulting to U.S. West, NASA, and the Swiss Bank Corporation. Dr. Sallach served as Director of Social Science Research Computing from 1998-2003, where, along in conjunction with Nicholson Collier, he designed the architecture of the Repast agent simulation toolkit. Since 2003, he has been Associate Director of the Center for Complex Adaptive Agent Systems Simulation at the Argonne National Laboratory, and also a Senior Fellow at the Computation Institute at the University of Chicago and Argonne National Laboratory. Dr. Sallach specializes in applying agent modeling and simulation technologies within social science domains. He draws extensively upon social theory, translating it into trans-scale social models. He has published articles in *IEEE Intelligent Systems, the Social Science Computer Review*, the *Journal of Mathematical Sociology*, *Rational and Society*, and other professional journals.

Jean Scholtz has worked in the area of user-centered evaluation of technology for 20 years. She is semi-retired and currently works part-time for the Pacific Northwest National Laboratory, primarily in user-centered evaluation of visual analytic systems. As such she focuses on users in law enforcement and intelligence analysts. Previously, she worked for the National Institute of Standards and Technology (NIST) where she developed metrics and methodologies for user-centered evaluations of programs for the Defense Department and the Intelligence Community. She is a founder and co-chair of the Visual Analytics Science and Technology (VAST) Challenge which offers participants an opportunity to use their visual analytics systems to analyze synthetic data with embedded ground truth. Dr. Scholtz's work history includes positions at Bell Telephone Laboratories and Intel Corporation. She was also on the Computer Science faculty at Portland State

University, Portland, OR. Dr. Scholtz holds a PhD in Computer Science from the University of Nebraska. She is a member of ACM and IEEE and has over 60 publications in user-centered evaluation.

Timothy G. Trucano is currently a Senior Scientist in the technical staff at Sandia National Laboratories in the Optimization and Uncertainty Estimation Department. He received his PhD in mathematical physics from the University of New Mexico in 1980. He began work at Sandia in August of that year. For the first 15 years of his career at Sandia, Tim's work focused on research, development, and applications of computational shock wave physics. Over this period of time Tim worked on problems of shock hydrodynamics, equation of state and constitutive modeling, hypervelocity impact phenomena, nuclear weapons effects, and radiation-hydrodynamics in high energy density physics applications. Beginning in 1995, Tim's work began to focus on issues of computational science verification and validation (V&V), as well as the application of uncertainty quantification. Currently, he works on technical and programmatic challenges for the NNSA Advanced Simulation and Computing (ASC) V&V Program, as well as research in the application of uncertainty quantification in computational prediction and the accompanying decision environment. Most recently, Tim has also begun to study the role of V&V in the application of complex social models to national security problems.

Jessica Glicken Turnley is President of Galisteo Consulting Group, Inc., a consulting firm in Albuquerque, NM. She also holds an appointment as Senior Fellow, Joint Special Operations University, USSOCOM. Through her work with Galisteo Consulting Group, she has provided services in policy analysis and national security, strategic business planning, organizational development, corporate culture change, and economic development to a wide variety of clients in the public and private sector. As a Senior Fellow, she provides research, analysis and concept development of selected special operations issues, initiatives and emerging concepts, focusing on those related to organizational and cultural topics pertinent to our own forces and national security organizations, as well as the adversary. Dr. Turnley's work has included several projects directly related to computational social modeling. She has engaged in several efforts focused on the social and cultural aspects of homeland security, the assessment of the terrorist threat, and the critical assessment of our own national security complex. She has served on the Defense Intelligence Agency Advisory Board. She also is working in the area of the social study of science. Other projects include support to the EPA in the development of approaches to assess social, cultural, and economic impacts at Superfund sites, the establishment of a bi-national applied research laboratory with Mexico in conjunction with Sandia National Labs, and research in the cultural aspects of workplace safety. Dr. Turnley has participated in regional economic development efforts, and engaged in organizational audits and development projects for local organizations. She has also organized and run focus groups and workshops on a variety of topics and managed strategic planning processes for small and large corporations.

Michael Vlahos is Professor of Strategy at the United States Naval War College. His is the author of Fighting Identity: Sacred War and World Change, an analysis of how war — as sacred ritual — shapes collective identity: And what it means in culture to be human. His career includes service in the Navy, the Central Intelligence Agency, Johns Hopkins University, and the U.S. State Department. An historian-anthropologist of war, he focuses on the relationships between civilizations, and the creative syncretism that is at the heart of change in history. He appears and posts on Huffington, the National Journal, and the John Batchelor Radio program (WABC).

Alyson Wilson is an Associate Professor in the Department of Statistics at Iowa State University and a Scientist 5 in the Statistical Sciences Group at Los Alamos National Laboratory. She is a Fellow of the American Statistical Association and a recognized expert in statistical reliability, Bayesian methods, and the application of statistics to problems in defense and national security. Prior to joining Iowa State in 2008, Dr. Wilson was a Project Leader and Technical Lead for Department of Defense Programs in the Statistical Sciences Group at Los Alamos National Laboratory (1999-2008). Dr. Wilson has served on numerous national panels, including the National Academy of Sciences (NAS) Committee on Mathematical Foundation of Validation, Verification, and Uncertainty Quantification (2010-2011), the NAS Committee to Review the Testing of Body Armor Materials for Use by the U.S. Army (2010-2011), the NAS Oversight Committee for the Workshop on Industrial Methods for the Effective Test and Development of Defense Systems (2008-2011), the Sandia National Laboratories' Predictive Engineering Science Panel (2008-2013). In 2006, the chaired the American Statistical Association President's Task Force on Statistics in Defense and National Security. Dr. Wilson is the founder and past-chair of the American Statistical Association's Section on Statistics in Defense and National Security. She is the incoming Reviews Editor (2011-2013) for the *Journal of the American Statistical Association* and the *American Statistician*, and she is a member of the *Technometrics* Management Committee. In addition to numerous publications, Dr. Wilson recently co-authored a book, *Bayesian Reliability*, and has co-edited two other books, *Statistical Methods in Counterterrorism: Game Theory, Modeling, Syndromic Surveillance, and Biometric Authentication* and *Modern Statistical and Mathematical Methods in Reliability*. Dr. Wilson received her PhD in Statistics from Duke University, her M.S. in Statistics from Carnegie-Mellon University, and her B.A. in Mathematical Sciences from Rice University.