Defense Threat Reduction Agency
8725 John J. Kingman Road, MSC 6201
Fort Belvoir, VA 22060-6201

Advanced Systems and Concepts Office

FINAL REPORT

2007 011

Verification and Validation of Social Science Simulations:
Some Thoughts for Consideration in National Security Applications

July 13, 2007

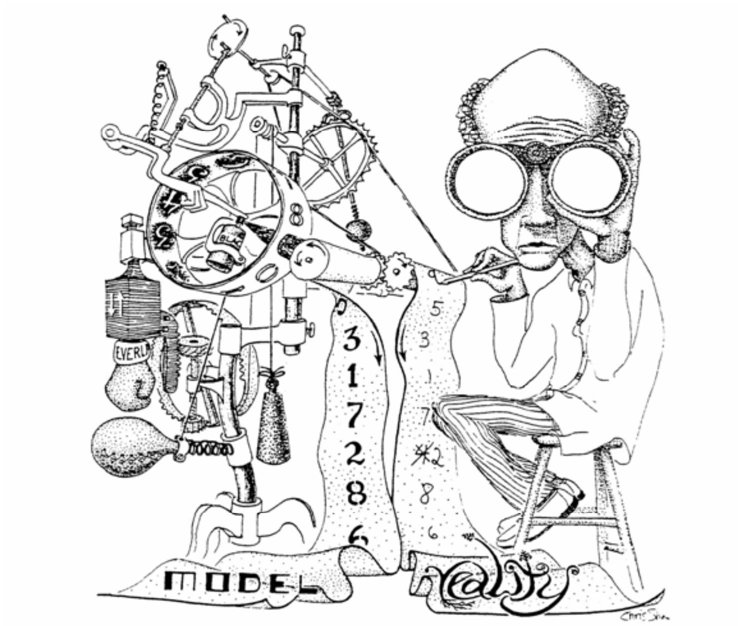
Author: Nancy Kay Hayden

Prepared by:
Sandia National Laboratories
PO Box 5800
Albuquerque, NM 87175

Contract Number 054137I

Verification and Validation of Social Science Simulations:

Some Thoughts for Consideration in National Security Applications



Nancy Kay Hayden
Sandia National Laboratories
Albuquerque, NM

July 13, 2007

Work performed for Defense Threat Reduction Agency/Advanced Systems and Concepts Office (DTRA/ASCO) Contract Number 054137I.

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.

Acknowledgements

Over the years, many colleagues have contributed to the ideas contained in this report. I especially thank those who supported and trained me in real time on methods for model validation and uncertainty quantification for decision analysis while working together on the Yucca Mountain and Waste Isolation Pilot Projects: Felton Bingham, Lyn Tyler, Jon Helton, Fred Mendenhall, Richard Lincoln, “Rip” Anderson, Mel Marietta, Martin Tierney, Steve Hora, and Walt Beyeler. More recently, I thank my friends and colleagues Jessica Glicken Turnley, Laura McNamara, and Wendell Jones for deep insights into the commonalities and differences between approaches to these problems in the social science, engineering, and information systems communities. Jessica deserves special thanks for original conception of the paper. I thank the program sponsors and colleagues at DTRA/ASCO -- Jennifer Perry and D.J. Patil in particular -- for insight, guidance, technical expertise and patience during the project. Finally, thanks go to Trey Hammond, Timothy Drummond, and William Prindle, who generously reviewed various versions of the manuscript and provided valuable review and editorial comments.

Introduction

Threats to national security have become increasingly complex and adaptive involving the intersection of technical, operational, and human considerations. Non-state actors that use non-traditional means of attack such as terrorism and the threat of acquisition or use of weapons of mass destruction (WMD) are of particular concern, as the numbers of actors, their potential repertoires of behaviors, and the dynamic, co-evolutionary nature of their actions with those of other actors compound the problem space enormously. The expansion along these several dimensions has encouraged the emergence of many new modeling approaches, as well as the innovative application of existing methods across domain spaces, to support transdisciplinary data collection, analysis, and decision support. The direct inclusion of social sciences in particular, and (although to a lesser extent) the humanities in the threat analysis endeavor introduces new concepts, frames of reference, and language to what historically was a field structured around the physical and engineering sciences.

Many of the new approaches with a growing advocacy in both the national security and social science communities derive from computational simulations of human behavior[1]. Within the national security community, these methods offer the potential to improve insights for situational understanding of conflict situations, the threats that arise from them, and options for responses to those threats. Within the social sciences community, there are some who see these methods as a complement to, or even a substitute for, traditional research methods, particularly for examining social phenomena derivative of organized complexity[2]. Richardi classifies three distinct cases where these approaches are most usefully applied[3]:

1. Numerical computation of analytic models –useful when a model is not analytically solvable for a relevant variable, when the model is stochastic, when the model can be solved for equilibrium but not for dynamic situations
2. Testing the robustness of analytic models – useful for investigating what happens on departure from some of the assumptions (e.g., structural homogeneity versus heterogeneity, bounded rationality)
3. Stand alone simulations of models of reality – useful for addressing problems that are analytically intractable or for which an analytic solution can not provide meaningful insights. In this case, analytic solutions of parts of the problem may exist, but not for the system as a whole.

[1] DTRA commissioned the Center for Nonproliferation Studies in 2002 to conduct a literature review of these approaches as they might apply to terrorist behavior modeling. The final report, "Literature Review of Existing Terrorist Behavior Modeling", (available at http://cns.miis.edu/research/terror.htm) revealed hundreds of new computational simulation models of human behavior developed in recent years.
[2] Goldspink, Chris (2002) "Methodological Implications Of Complex Systems Approaches to Sociality: Simulation as a Foundation for Knowledge", *Journal of Artificial Societies and Social Simulation* vol. 5, no. 1, http://jasss.soc.surrey.ac.uk/5/1/3.html
[3] Richardi, Matteo, (2003). "On the Use of Agent-Based Simulations", Laboratorio Riccardo Revelli Center for Employment Studies, Torino, Italy.

These applications hold promise for systems-level analyses in which the dynamic interactions between complex natural (physical) and human systems are paramount[4]. Indeed, though some members of the community are embracing these approaches to as means to overcome challenges of analyzing dynamic social processes, there is some resistance – especially among some of the more qualitative-focused scientists and practitioners -- to these approaches. One of the key issues of debate within these communities is the verification and validation (V&V) of these models, especially in decision making environments. This includes the methods by which V&V can be conducted and the extent to which it is possible.

Definitions of verification and validation themselves are an ongoing source of debate. For the purposes of this paper, they are defined as follows:

- Verification is a demonstrable and repeatable assurance that the computational representation of a conceptual model, the input to it, and the representation of the output, is accurately calculated through the algorithms and simulation processes employed to meet the design criteria.
- Validation is the assurance that the simulation as an entirety (data, conceptual model, computational representation, simulation, and output) is a "true" representation of the reality being approximated.

In short, validation shows that you built the right model, and verification shows that you built the model right.

As discussed by Turnley, the requisite verification and validation process for social science simulation models parallels, in the abstract, the process required for computational models in the natural sciences[5]. In both cases, one begins with observation of data about elements of the real world and then invokes or formulates a "conceptual model" of how those elements behave in, and interact with, the real world. This conceptual model is then formalized into computational form, which is then rendered into "code". The code is implemented in some type of simulation process to generate output, which is then analyzed and interpreted to develop conclusions. (See Figure 1.)

Data → **Conceptual Model** → **Computer Model** → **Simulation Process** → **Output**

Figure 1. General steps in verification and validation of computational simulations

---

[4] Hayden, Nancy Kay, (2007), "The Complexity of Terrorism: Social and Behavioral Understanding – Trends for the Future", *Mapping Terrorism Research: State of the art, gaps, and future direction,* Magnus Ranstorp (ed), Routledge, NY, NY. pp. 292-315.

[5] Turnley, Jessica Glicken, "Validation Issues in Computational Social Simulations", working paper, Galisteo Consulting Group Inc., May 2005.

In simple linear systems, the input data and generated output can compared to some standard (real world, other codes, natural analogues, etc.) to ensure that the entire process is a faithful representation of reality.

However, this is not the case for complex, nonlinear dynamical systems. Questions must be addressed *at each step*, about the ***validity*** of both the *content* and the *procedural way* in which the step is conducted as well as ***verification*** of the ***mathematical accuracy*** of the method chosen to carry out that procedure. Questions about mathematical accuracy, though nontrivial, are the easier of the two sets of questions to address, and depend primarily upon correct implementation of the computational procedure chosen. Questions about validity of the data, the conceptual model and its underlying theoretical bases, and the computational instantiation of that model are more difficult to formulate as well as address, as the questions themselves are dependent upon the epistemological context of the analysis for which the model is being used. Additional complexity is added when multiple interacting models are invoked, involving a mix of natural and human systems with different behavior timescales and data density.

This paper discusses 1) considerations of analytic framing as a precursor to the validation process for complex social systems, 2) perspectives from existing social science and artificial intelligence literature on validation of computational models for simulating social behaviors, and 3) perspectives of users and decision makers to highlight concerns and to ensure that the proposed framework is accepted by the target communities. Together, these considerations can provide the beginnings of a framework for assessing conditions under which computational simulations of social phenomena are desirable, which types of simulations are most appropriate, and approaches for validation. Further work will be necessary to develop a robust framework.

Considerations of Analytic Framing for Complex Systems

The crux of this paper is to present some key thoughts on verification and validation of computational simulations of complex, self-organizing social systems. However, before considering validation of such analytic models, one must first be able to show that the methodology inherent to the model is appropriate for the problem being addressed, and has been well grounded in terms of three critical dimensions: the epistemology of the analysis (e.g., what knowledge does the analysis purport to achieve and how is it to be used?), the degree of system complexity to be considered, and the timescale necessary to capture the system dynamics (Figure 2). This section discusses this framework.

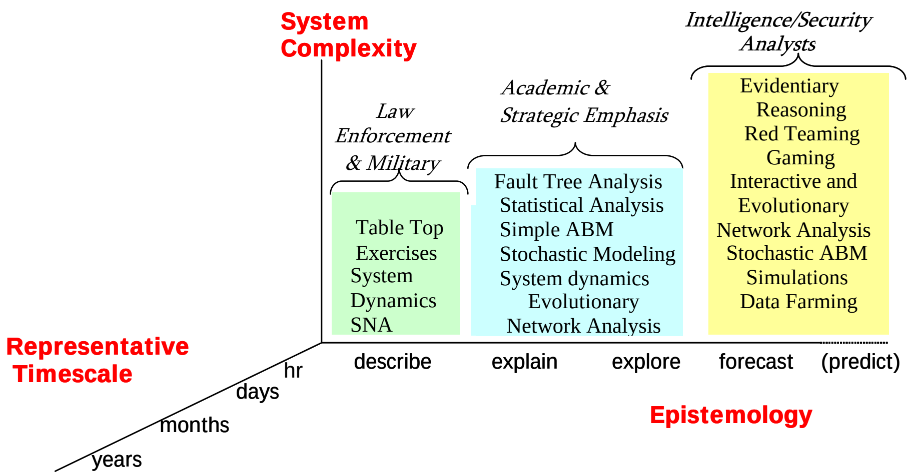


Figure 2. Analysis framework for complex systems simulations

As discussed in previous work, epistemological characterization is an essential precursor to validate complex systems models, as it is fundamental to holistically choosing the appropriate analysis method, data, and timeframe of relevance[6]. However, this critical first step is often overlooked rendering subsequent validation efforts difficult at best. Analytic inquiry can range from seeking to understand *what is*, to understanding *why something is*, to seeking *what is possible*, or predicting *what will be*. Representative (but neither prescriptive nor comprehensive) epistemological regimes are shown in Figure 2, along with typical analytic methods used within those regimes and typical communities of analysts who employ them.

On one extreme, where one seeks to answer the question of *what is*, the dynamic, iterative, and interactive nature of complex systems requires that analysis be grounded within an appropriate, finite timeframe and a specific context. Law enforcement and front-line military (expeditionary) forces provide examples of communities with a primary focus on answering questions about what has just happened or is currently happening. These communities require frameworks and analytic methods that facilitate understanding of direct observations in relatively short timeframes.

On the other extreme, some analytic epistemologies seek to *predict what will be*. However, as discussed below, a characteristic of complex systems is their unpredictability under dynamic conditions. Therefore, attempts to quantitatively and explicitly predict complex system behaviors will fail. Unfortunately, the environment within which much of national security analysis occurs fosters an approach of attempting to achieve such "point" predictions. Most analysts will caution that in this regime, the most one should attempt to generate is insight into the dynamics that might shape future behaviors, as opposed to conclusive predictions.

[6] Hayden (2007)

Questions about systems in a state of organized complexity are intermediate between these two extremes, and require appropriate analytic methods. There is a need to map analytic methods to epistemological regimes, and to understand the interfaces between these analytic methods in order to better understand best practices for verification and validation. The barriers to achieving this are numerous – such as the differential level of system complexity that different methods have been developed to represent, the access to distributed data and differentiated classification of information that is required, the different cultures, values, and practices of the analytic communities, and different ontological representation of concepts that are "named" differently in different problem-solving domains[7].

Different players play different roles in to the verification and validation process across these technical domains. Key players are the analysts who have problems in some defined portion of the real world that is of interest, subject matter experts who know something about that portion of the real world and who create hypotheses and generate theories and data to support them, simulation model developers who can formalize those theories and data (knowledge) in simulation models, and decision makers who evaluate and make judgments about the real world on the basis of results. In the end, the decision maker must be convinced of the validity of models, the integrity of the process by which they have been used to represent some aspect of the real world, and the quantification of uncertainty in results in order to take action.

These latter points bring us directly to the second and third dimensions that must be addressed in building an analytic framework: identifying the level of system complexity and specifying the timeframe of relevance. Let us consider system complexity first. Multiple definitions have been proposed from various perspectives (e.g., probabilistic, computational, and algorithmic). A common element across all measures of complexity is that they are based on characterizing systems on a spectrum between order and chaos[8] (or randomness).

Varying definitions of complexity and associated measures have been proposed by the scientific community from a diversity of perspectives. Probabilistic, computational, and algorithmic definitions involve measuring how much information (or information processing) is required to completely describe a system and reduce uncertainty[9]. Crutchfield has shown that these formulations, which describe deterministic complexity

[7] For example, the concepts of value transference, social contagion, and epidemiological spread of disease are conceptually similar from a modeling perspective but are ontologically and mechanistically represented differently in their particular domains.

[8] Chaotic systems are nonlinear dynamical systems that under specific conditions exhibit dynamics that are sensitive to initial conditions (popularly referred to as the butterfly effect). As a result of this sensitivity, the behavior of chaotic systems appears to be random at early times, because of an exponential growth of errors in the initial conditions. This happens even though these systems are deterministic in the sense that their future dynamics are well defined by their initial conditions (although not known a priori), and there are no random elements involved. This behavior is known as deterministic chaos, or simply *chaos*.

[9] Gell-Mann, M. (1994) *The Quark and the Jaguar - Adventures in the Simple and the Complex*. New York, NY: W. H. Freeman and Company. ; Holland, J. (1998) *Emergence: From Chaos to Order*. Cambridge, MA: Perseus Books. Suh, N. (2005) *Complexity: Theory and Applications*. New York, NY: Oxford University Press.

in terms of randomness, are insufficient for describing the structural complexity present in most natural systems[10]. He resolved this by the introduction of a statistical metric of complexity based on stochastic processes and entropy that are correlates to a relative measure of structure. In his formulation, relative measures of both randomness and structure are necessary for determining a system's complexity, as shown in Figure 3.

This structural measure proposed by Crutchfield is also more appropriate for the social and behavioral sciences, as it correlates to varying degrees of organization – or structure, regularity, symmetry and intricacy – in a system's behavior and/or architecture[11].

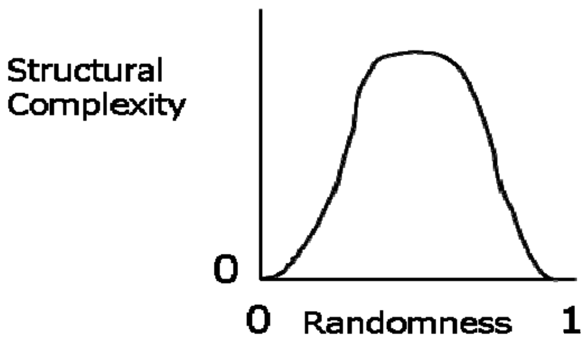


Figure 3. Statistical complexity is based on the notion that randomness is statistically simple: an ideal random process has zero statistical complexity. At the other end of the spectrum, simple periodic processes have low statistical complexity. Complex processes arise between these extremes and are an amalgam of predictable and stochastic mechanisms. Reprinted with permission.

At the extremes of randomness, the system is structurally simple. Statistical complexity – which is correlated to structure – is greatest in the intermediate regime. In complexity literature, this intermediate regime is referred to as the "edge of chaos", and is where some of the most interesting system behaviors occur - such as surprise, innovation, and phase transitions. It is at phase transitions, for example, that the forces leading to order and disorder compete, producing unique critical states, more complex than those away from the transition[12]. However, in terms of V&V, this is the hardest region to address; is it is the region in which emergent behaviors are manifested.

Crutchfield's complexity-entropy formulation has significant implications about the ***predictability*** of complex social systems. At the extremes of randomness, where statistical (stochastic) complexity is lowest, predictability is greatest. This is true even in deterministic chaotic systems in the short term, although long term predictability of such systems is impossible[13]. However, these predictions are of limited value as they represent average behaviors, whereas the more significant behaviors are those of the outliers.

---

[10] Crutchfield, J. (2003) 'When Evolution is Revolution: Origins of Innovation' in J. P. Crutchfield and P. Schuster (ed.) *Evolutionary Dynamics: Exploring the Interplay of Selection, Neutrality, Accident, and Function*. New York, NY: Oxford University Press.
[11] Crutchfield, J. (1994) *The Calculi of Emergence: Computation, Dynamics, and Induction.* Santa Fe Institute Working Paper 94-03-016.
[12] Holland, J. (1998) *Emergence: From Chaos to Order*. Cambridge, MA: Perseus Books.
[13] Flake, G. W. (1998) *The Computational Beauty of Nature: Computer Explorations of Fractals, Chaos, Complex Systems, and Adaptation*. Cambridge, MA: The MIT Press.

The implications for validating models of complex systems are profound. It means that at the very outset, one must ask the question, “What model, for what purpose?”, before going on to try to show that the model accurately represents the processes in the real world situation that one is attempting to address. For example, for cases that fall into regimes that are inherently unpredictable (due to complexity), we must cease to ask questions about *explicitly predicting* future events (e.g., *How likely is it for an event of type x to occur over time period y perpetrated by group z at location a?*) and instead ask questions that seek to *explore possibilities* of future behaviors and the key indicators for those behaviors in terms of dynamic patterns of interactions and the underlying structures upon which those transactions take place (e.g., *Given that we see certain types of relationships forming and interactions taking place in certain contexts, what are the risks of a terrorist event?*). The model must be chosen accordingly in order to validate it for application to these cases.

The third dimension to consider in framing a valid social analysis using modeling and simulation relates to the timescales of behavior changes in the system being examined. These changes can be brought about through **evolution** and/or **innovation**. The dynamics of evolution and the many mechanisms through which it occurs - e.g., *selection, variation, learning, adaptation* – explain the diversity of structure, function, and behavior in complex systems. Innovation – which can be incremental, radical, or systemic – is specific to the introduction and transfer of something new within the system in order to solve a problem. Together, these explain why complex systems that stay complex are not in a steady-state of equilibrium – but instead exist in dynamic, quasi-stable non-equilibrium states. Two fundamentally different types of evolutionary processes drive change: those that occur as a result of differentiation between individual entities in the system, and those that occur as a result of all entities simultaneously undergoing a similar, generalized transformational process[14].

Any or all of these processes may be operative in a complex system, depending on both intrinsic properties and external conditions. The modeling approach must be carefully selected to accommodate these appropriately. If the system contains elements with greatly differing timeframes of behavioral changes, both verification and validation are more difficult at all steps in Figure 1.

Perspectives from Social Sciences and Artificial Intelligence Literature

The preceding discussion about simulating behaviors in complex systems sets the stage for considering the crux of this paper, namely,

1. What are some issues around validation of computational methods for analysis of these behaviors within social systems?
2. What are some approaches that address these issues?

[14] Lewontin, R. (2000) “Evolution”. in Bar-Yam (ed.) *Unifying Themes in Complex Systems (I)*. Boulder, CO: Westview Press.

3. What concerns need to be addressed when used in national security applications?

The following section draws on literature in both the social sciences and artificial intelligence, and presents a range of perspectives on these topics. First, epistemological considerations, introduced earlier, are discussed with an emphasis on what they mean for validation specifically. Second, meta-level validation approaches are presented that are applicable to the entire analysis space shown earlier in Figure 2. Finally, some specific techniques are highlighted as examples of how different validation methods can be targeted towards specific types of models and regions of the analysis space.

Epistemological Considerations

In the previously cited review article for the *Journal of Artificial Societies and Social Simulation*, Goldspink discusses reasons for advocating the use of computational simulations in the social sciences, the specific advantages of the method for studying such phenomena, as well as the limitations and problems that need to be addressed for the method is to gain wide acceptance[15]. He points out that the phenomena arising from complex organization and non-linear interactions are being found to be increasingly important in the social, cognitive, behavioral and organizational sciences. This growing realization challenges existing social science research methods, where the systems involved display emergent behavior. Ilgen and Hulin[16] state, for example,

> "Our methods and theories remain far better suited for the deterministic and linear corners of [organization science] than for the well populated chaotic regions of it."

Social scientists consider simulation methods to be particularly useful when:[17]

> "The phenomena to be studied are either not directly accessible or difficult to observe directly."

Compared with standard research approaches which involve observation of, or experimentation with, people within organizational contexts, Prietula et al[18] also note that:

> "Computational models are generally less noisy, easier to control, more flexible, more objective, and can be used to examine a larger variety of factors within less time."

---

[15] Goldspink (2002)
[16] Ilgen, D.R & C. L. Hulin (eds) (2000), "Computational Modeling of Behavior Organizations: The Third Scientific Discipline", American Psychological Association, Washington DC.
[17] Gilbert, N. (1996), "Computer Simulation of Social Processes", Social Research Update, Issue Six, http://www.soc.surrey.ac.uk/sru/SRU6.html.
[18] Prietula, M.J., K.M. Carley, and L. Gasser (eds.) (1998), *Simulating Organizations: Computational Models of Institutions and Groups*, AAAI Press.

In addition to emergent structures, there is growing interest among some sectors of the social sciences community in decentralized phenomena and self-organization within dynamic, nonlinear complex systems. It is in this context, Goldspink argues, that simulation methods are attractive as adjuncts to, or substitutes for, conventional methods and for the further testing or development of social theory. Simulations used in this way can contribute meaningfully to predictive analysis, proof, and discovery[19], leading to 'a new way of conducting science' - one which bridges traditional inductive and deductive approaches[20]. Starting with a set of assumptions, simulation generates data that can be analyzed inductively. Unlike typical induction, however, simulated data comes from a specified set of rules rather than direct measurement of the real world, which has implications for validation. Goldspink advocates use of simulation within an overall framework that incorporates computational techniques in a broader methodological mix. These techniques include use of historical narrative and situated research methods (increasingly accepted as post-positivist methods) for capturing dynamic patterns in real world systems. Such a mix has the potential, he argues, to harness the strengths of the simulation methods while offsetting some of their weaknesses.

Validation approaches

Validation approaches for computational models and simulations of logical reasoning and human behaviors are an ongoing area of research in the artificial intelligence and social science fields. The literature has produced a variety of concepts useful for developing a validation framework. Some of the relevant citations for each of the steps in Figure 1 are discussed below.

Stages of Model Validation

Zeigler has proposed three stages of model validation (for both natural and social systems) that mirror the epistemological analysis framework proposed by Hayden[21]:

1. Replicative validity: the model matches data already acquired from the real system (retrodiction);
2. Predictive validity: the model is run "blind" of the data from the real system (because the data has not yet been obtained, the simulation is for a future event, etc). Upon acquisition of data, the model matches the real system;
3. Structural validity: the model "not only reproduces the observed real system behavior, but truly reflects the way in which the real system operates to produce this behavior."

---

[19] In this discussion, *discovery* refers to simulation's potential to reveal the unexpected, to expose unanticipated relationships or results by making explicit what was hidden in the implicit 'rules' of operation or characteristics of components.

[20] Axelrod, R. (1997), "Advancing the Art of Simulation in the Social Sciences", *Complexity*, Vol. 3, No 2, John Wiley N.Y. pp. 16- 22

[21] Zeigler, B. P. (1985) *Theory of Modeling and Simulation*. Krieger, Malabar. (Reprint, first published in 1976, Wiley, New York).

Zeigler points out that a key difference in validation models of natural and social science systems is that data collection is often much more expensive and difficult - indeed at times impossible - in the case of the latter. One of the reasons for this is the changing nature of social processes. Data collected at one point in a social process cannot represent data at all points in the social process. In addition, the structural variations possible in natural systems are often much more limited and targeted than for social systems, such that a structurally valid model can be easier to find in the case of the former.

Turnley has noted additional differences between models developed for physical and social systems, namely[22]:

- In many cases, modeling in physics is really math, not physics
- Conceptual models of physical systems are (comparatively) unchallenged
- Physical problems can be 'binned' by the way the problems are defined
- Even when stochastic, physical phenomena are 'essentialist' in nature (one atom is like another atom) and can be directly represented as equations and "code"
- Simplifying assumptions in physical systems are often able to be captured in primitive equations
- Population thinking about biological and human systems says there is no 'type'; each individual is unique
- The nature of individuals change as they move through time and space
- In social systems, 'mean values are abstractions', without direct physical meaning

A final consideration is that the ability to "prove" the validity of the underlying social science theory tends to be more problematic than in the case of many natural science theories, thus impacting the validation process. Oftentimes what might be considered "valid" for one theoretical perspective (i.e. provides evidentiary support for a particular view on real world human behavior) may not be considered valid for another theoretical perspective in the social sciences. It is important that these differences are explored in developing approaches to model validation.

Approaches to Validation

While there is no consensus for any ONE right approach to validating models representing social processes, the literature is consistent about the limitations of retrodictive and positivist approaches. For example, McKelvey argues that axiomatic approaches typical of positivist methods of validation (typical in natural sciences) rarely work in the social sciences, particularly when the phenomena being studied is

[22] Turnley, Jessica Glicken, Galisteo Consulting, Inc., Personal Communication, October 2006

emergent[23]. Consequently, he argues for a semantic approach, in which theory and model are viewed independently. Truth testing takes place in two ways[24]:

> "Firstly, experiments are conducted using the model. This allows the prediction of theory to be tested in a controllable way, albeit in a simplified analog of the real phenomena. Secondly, ontological adequacy is tested by comparing the isomorphism of the models idealized structures/processes against that portion of the total real-world phenomena defined as within scope of the theory. Alternative theoretical conceptions may be used to understand any model, and alternative models derived from any given theory."

Rules of Validation

Independent of the approach to validation, Leik and Meeker[25] suggest a set of rules for building social simulations which have some hope of being validated:

1. Every tie from the simulation to the model to the substantive theory needs to be made explicit;
2. The way each algorithm in the simulation works needs to be laid out so others can judge its appropriateness;
3. Every constraint on variable, parameters, numbers of runs and so forth needs to be justified;
4. Every decision about what to examine and report needs to be made explicit;
5. All justifications must be in light of the substantive interpretations to be made of the model being simulated.

This last point in particular ties back to the need for defining the epistemological framework of the analysis to begin with, as discussed in the first section. When choosing a simulation model, one must evaluate how well the interpretations to be made from the results (i.e., questions to be answered) can be defended on the basis of the model itself. For example, if conducting an analysis to explain a certain phenomena, are there multiple pathways possible through the model to reach the same end state? If so, any one particular explanatory interpretation may be difficult to justify. A stochastic process, with an average set of results and determination of significance of outliers might be required as part of the process in this case.

Simulations and Observation, Data, and Theory

---

[23] McKelvey, B., (1997), "Quasi-Natural Organization Science", *Organization Science*, Vol 8 No. 4, pp. 351- 380.
[24] McKelvey, B. (1999), "Complexity Theory in Organization Science: Seizing the Promise or Becoming a Fad?", *Emergence*, Vol 1 No 1., pp. 5-32.
[25] Leik, R. K. & B. F. Meeker. (1995) "Computer Simulation for Exploring Theories: Models of Interpersonal Cooperation and Competition", *Sociological Perspectives*, Vol. 38, No. 4, pp. 463-482

Troitzsch addresses the problems with both replicative and predictive validity that stem from the relationship between data and theory. Inspired by an extensive exchange within an on-line simulation community on the role of simulation in theory building in the social, economic, and management sciences, Troitzsch developed a paper reviewing arguments on the use of quantitative and qualitative computational models to build theoretical explanations for relationships between observations and data and providing an overview of topics in validation from a structuralist point of view[26]. In the paper, he points out the fundamental difficulty in distinguishing between observation, data, and theory at the outset of building a model:

> *"Being aware that observation (as contrasted to just looking around in the world) presupposes at least some primitive form of theory (which tells us which entities and which of its properties to observe and which relations between them to register to find out whether there are some regularities), we should admit that our assumptions and our observation are not independent from each other. And we should admit that in most cases computational (and other) models do not directly start from observation data but from a theory which in turn should build on, but often does not refer explicitly to observation data. Instead, we often start from a verbal theory which expresses our (or other authors') belief in how reality works, comparing simulation results with stylized facts instead of observation data."*

Citing the work of Zeigler on the stages of model validation, Troitzch notes that since data are often very poor in the social sciences, early models tried to be structurally valid and did not bother much about replicative or predictive validity.

As an example, Troitzsch cites Sugarscape[27], where the question "Can you explain it?" is replaced by the question, "Can you grow it?" (where the term "it" refers to an observation). This leads to "explanation" of macrostructures based on the generative capabilities of simulated microstructures. In and of itself, this may or may not be valid, and demands another approach for model validation. (See discussion below on the Alignment method proposed by Axtell et al).

On the other hand, one may use micro-analytical simulations based on a large collection of quantitative observations to generate predictions of qualitative structure formations (e.g., kinship networks, neighborhood transitions, etc) in the future. In either case, Troitzsch argues that computer simulation programs can be seen as models of theories from the point of view of the structuralist program in the philosophy of science. In his view, this means that computer simulation should always have an empirical claim. This claim can come in different forms—from quantitative predictions of future measurements to qualitative descriptions of possible scenarios. Both can be used to validate the theory behind the simulation model.

---

[26] Troitzsch, Klaus G. (2004) "Validating Simulation Models", Proceedings 18th European Simulation Multi-conference, Graham Horton SCS Europe, ISBN 3-936150-35-4.

[27] Sugarscape is a "bottoms-up" agent based model developed by Joshua Epstein and Robert Axtell at Brookings Institute as a research tool for studying aggregate effects of interactions among agents, where the interactions are governed by a few simple rules. The model has been used extensively as a research tool to study a variety of complex situations, such as cultural evolution, warfare, and population dynamics. See Sugarscape webpage: http://www.brook.edu/es/dynamics/sugarscape/default.htm

However, these arguments fall short when social theory is lacking – especially when that is due to the lack of empirical evidence upon which to base a theory. Indeed, it is in this instance – which corresponds to Richard's third case - that some argue for social simulations to serve as substitutes for "empirical claims" by generating simulated data to inspire and test new social theories. This in turn calls for fundamentally different approaches to validation.

Alignment Method

Pioneers in the development of computational methods (e.g., agent-based modeling) for simulating human behavior, Robert Axtell, Robert Axelrod, and Joshua Epstein, along with Michael Cohen, pioneered some of the first systematic academic thought on verification and validation issues associated with the way in which data and conceptual models are instantiated in computational simulations. In 1995, through a process of close comparison of models, they advanced the concepts and methods for determining "domain of validity" that typically can be obtained for mathematized theories. This process, which they called "alignment of computational models" or "docking", determines whether two computational models can produce the same results; this in turn became their basis for critical experiments and for tests of whether one model could subsume another[28]. They illustrated the concepts and methods using as a target a model of cultural transmission built by Axelrod, and compared that to results from the Sugarscape model developed by Epstein and Axtell. Their approach is widely cited in the literature and is suitable in cases that do not involve too many entities in interdependent dynamic relationships. However, for very complex analyses, where it may take years to develop the appropriate conceptual relationship representations between many entities of different types and timeframes of behaviors and instantiate these into a computer model, it can be impractical to implement.

Constraint logic programming

As shown in Figure 1, examining the simulation process itself (that is, how the computer model is implemented) is a key part of validation. Simulation processes for exploration and predictive analyses of large multi-agent systems often involve many iterations from random starting points to generate a large envelop of possible trajectories of behaviors. Typically, a huge amount of computation is required in these cases when experimenting with factors bearing on the dynamics of a simulation to tease out what affects the shape of this envelope.

---

[28] Axtell, Robert, Robert Axelrod, Joshua M. Epstein, and Michael D. Cohen, (1995) "Aligning Simulation models: A Case Study and Results", *Computational Mathematical Organization Theory*, 1(2), pp. 123-141, http://citeseer.ist.psu.edu/axtell96aligning.html

The problem of validating these types of simulations is addressed by Teran et al[29]. Their proposed method systematically explores this envelope "using search techniques for tendencies and proving their necessity relative to a range of parameterizations of the model and agents' choices, and to the logic of the simulation language. The exploration consists of a forward chaining generation of the trajectories associated to and constrained by such a range of parameterizations and choices." While constraint search is not something new in declarative programming -- both backward and forward chaining inference based tools are in use for constraint reasoning -- their work is the first application of this method to simulations in large multi-agent systems based on parameters of the model and the possible behavior 'trajectories'. In particular, it allows for a validation approach when searching for *emergent* behaviors in the system. As pointed out by Edmonds, this is particularly important in social simulation, where it is important to search for behaviors that are not path-dependent (i.e., *non-contingent*), and *non-expected* (i.e., emergent) from the simulation design[30].

Rather than validate dynamics of simulations through post-hoc analysis, a systematic and controlled exploration of the simulation dynamics is conducted through internal programming constraints, subspaces of trajectories are analyzed. The method uses a modular structure according to strategic parts of a simulation[31]:

> "A first module, *model*, sets up the *static structure* of the simulation; then a second module, *prover*, generates the *dynamics* of the simulation; and finally a *meta-module* is responsible for *controlling* the dynamics of the simulation. The second characteristic of this platform is a *partitioning* of the space of rules and *splitting of transition rules* by strategic time intervals, parameters and choices…Constrains are context-dependent (over the semantic of the trajectory itself) as the meta-module is *able to access the semantics* of the simulation setting up in advance one among the possible combination of agents' choices and model parameters for each run."

By so doing, the simulation has "built-in proof" of the validity of qualitative results in terms of the shape of the envelope of possible behavior trajectories. The authors contend that through controlled relaxation of constraints, the method can also be used for hypothesis testing.

Compositional Verification Approach

While the primary focus of this paper is on validation approaches, a particularly interesting verification approach by Jonker and Truer for large multi-agent systems bears

[29] Teran, Oswaldo, Bruce Edmonds, and Steve Wallis, (2000) "Constraint Exploration and Envelope of Simulation Trajectories", First Workshop on Rule-Based Constraint Reasoning and Programming at the First International Conference on Computational Logic (CL2000), July 24-28, 2000, Imperial College, London, UK., http://citeseer.ist.psu.edu/335166.html
[30] Edmonds, B. (1999), **"**Modeling Bounded Rationality In Agent-Based Simulations using the Evolution of Mental Models". In Brenner, T. (Ed.), *Computational Techniques for Modeling Learning in Economics*, Kluwer, pp. 305-332.
[31] Teran et al (2000).

mentioning. For such systems the boundary between verification and validation becomes quite "gray". In Jonker and Truer's approach, which has some inherent validation logic built into it, verification is defined to be the "proof that, under a certain set of assumptions, a system will adhere to a certain set of properties, for example the design requirements".[32]

Their approach uses mathematical proofs in distributed, formal analyses of relations between properties and assumptions to demonstrate that the system -- together with the assumptions -- implies the properties that it needs to fulfill. They propose a compositional method based on well-structured proofs of the process abstractions at different levels that are operative in both static and dynamic agents, with variants of pro-activeness and reactiveness. Formalized in terms of temporal semantics, re-useable requirements are derived from properties of agent components and their interactions with the "world" through a "task" models. The approach has been successfully applied in the multi-agent system DESIRE, using generic broker agent architecture[33].

Perspectives of Users and Decision Makers

The previous sections looked at issues around, and approaches for, verification and validation of complex social simulations from the perspectives of complexity science, social sciences, artificial intelligence, and the modeling community. In addition to model developers and users, perspectives of the ultimate decision maker[34]s, are equally important to consider in evaluating whether or not to use these types of models, and when using them to develop a validation processes to confidently apply them to assess problems of national security[35]. The process of interpreting model results, establishing the degree of confidence in them, and making recommendations for action on the basis of them is best done collaboratively between users and decision makers, but often falls to either one or the other.

One of the first things to ask as a user and/or decision maker is, "How will the model help me solve my problem?" In some cases, the answer will be different depending on the point of view. This should impact the overall simulation design process to assure validity of model for the intended purpose, as discussed in the introductory sections. For example, a user's requirements might be around assessing relative strengths of evidence for competing hypothesis to explain behaviors, whereas a decision maker's requirements

---

[32] Jonker, Catholijn M. and Jan Treur (1997) "Compositional Verification of Multi-Agent Systems: a Formal Analysis of Pro-activeness and Reactiveness", Lecture notes in *Computer Science*; Vol 1536, pp. 350-380.

[33] Jonker, Catholijn M. and Jan Treur (2002) "Compositional Design and Maintenance of Broker Agents", in *Intelligent Agents and Their Applications*, ISBN 3-7908-1469-5, Physica-Verlag GmbH, Heidelberg, Germany, pp. 149-172.

[34] In this discussion, a user is assumed to be an analyst or modeler whose job it is to represent reality through the model in order to answer questions about a particular problem; a decision maker is one whose job it is to take action on the basis of the results that come out of simulations using the model.

[35] The discussion in the section draws on the work of Jessica Glicken Turnley, Galisteo Consulting Inc., personal communication, October 2006.

might be around what actions should be taken that assures the best options are followed in either case.

Both points of view – user and decision-maker -- should drive model and simulation design requirements. As noted previously, in some cases a model can be a theory-building tool. In others, it can be used to'create' a prediction or representations of possible 'future states'. In yet others, models can be used in decision-making situations as mediators to gain further insights into a situation and to help in conceptualizing alternative views of possible future states. In this sense, models can advance the processes of learning about a problem, but do not necessarily contribute directly to the solution of the problem. Models can also be expressions of problems themselves, whereby in the development of the model, a consensus can be obtained among a diverse group of stakeholders[36].

As mentioned previously, a validation process must provide confidence that the simulation model is an accurate representation of the problem and captures the dynamics important to resolving the problem from the perspectives of both users and decision makers. The degree of confidence necessary will depend on the risk nature of the decision maker and may vary according to consequence of the decision, time frame available to make the decision, other information sources available to the decision-maker etc. As this risk factor (be it qualitative or quantitative), is difficult to measure or assess, it is ideal if the ultimate user and/or decision maker and their requirements are an integral part of the model building and validation process. This also enhances learning, as model construction is often the time of greatest learning about the relationships between elements, the factors that influence those, and how both co-evolve in time. As stated by Moss,

> "Good validation of social simulation requires prediction but a good prediction is not always a sufficient indicator for validity….Descriptiveness is also a good indicator…for the validity of ... models… When we model social processes in a participatory context, then agreement of the participating stakeholders on the validity of the model can be a reasonable indicator for the validity of the model."[37]

Another key to increasing a user and/or decision-maker's confidence in the model is to establish a clear communication structure between the modeler, user, and the decision

[36] For example, collaborative model building to achieve consensus is useful when there are equally likely explanations for phenomena that cannot be proved or disproved, or oppositional possible future states that impact exogenous conditions in the model. In these cases, the impact of adopting different assumptions can be explored and, in the best of all cases, agreement can be reached on what differences do NOT matter, and which ones do. Where significant differences are shown to matter, all parties can agree on courses of action to account for uncertainties introduced by these differences. This was the case in the use of models to define the scientific research program necessary to support the license application for the Waste Isolation Pilot Plant, as the world's first (and only) licensed geologic repository for nuclear waste. See reference to Prindle et al, 1996.

[37] Moss, S. (2001) Editorial introduction: Messy Systems — the Target for Multi Agent-Based Simulation. In S. Moss and P. Davidson (eds), *Multi-Agent-Based Simulation*, pp. 1–14. Springer, Berlin.

maker. Within this structure, there should be a well-articulated processes for communicating uncertainty, recognizing that decision makers tend to want a ‘number’ or a ‘bright line’ for action out of the model. This implies that there has been agreement on the important output metrics for making decisions with the model, how to assess the uncertainty of that output (which includes, but is not limited to, uncertainty in the model itself) and some consensus around knowing when enough modeling has been done to reach the level of confidence desired. Ideally, this communication structure is established at the outset of the model design process and maintained throughout its implementation and analysis of results for the reasons mentioned previously in this section.

Questions about uncertainty need to be differentiated from questions about validation in models, as there are many sources of uncertainty beyond those involved in the validation of a model. Methods for uncertainty quantification try to pinpoint what these sources are and how to reduce the uncertainty, once the sources are known. Validating models is only one of the means by which uncertainty is addressed. Gathering additional information (through ongoing research, experimentation, or simulations, for example) might also be valuable to address uncertainty. It is important to decision makers to understand to what degree uncertainty can be reduced by model validation and what uncertainty will still remain.

In the nuclear science and engineering communities – e.g., weapons, reactor safety, and waste management -- the stakes have been quite high for decisions made on the basis of model simulations; significant research dollars and efforts have accordingly been invested into methodologies for the iterative processes of model verification and validation, uncertainty quantification, information gathering (to reduce uncertainty), and risk assessments[38]. Although these models are primarily (but not entirely) based on physical (as compared to human/social) systems, these systems have high degrees of residual uncertainty which must be understood in the context of decision-making risk. Some of the sources of uncertainty bear similarity those in simulating social systems: namely, the dynamic, changing nature of the systems over the time-frames being modeled; the impact of unknown future events on assumptions in the models; and the absence of data in some modeling regimes that are outside of original design specifications. Accordingly, some of the approaches developed for better risk-based decision making that make explicit these uncertainties and the risk they pose to the decision-maker in the nuclear science and engineering programs bear investigation for application to social simulations. Two particular examples where methods have been successfully applied are those used to quantify uncertainty in the annual recertification by

[38] For recommendations of design of verification and validation benchmarks based on lessons learned in these applications, see Oberkampf, W.L. and T.G. Trucano, “Verification and Validation Benchmarks”, SAND 2007-0853, Sandia National Laboratories, Albuquerque, NM, to be published in Nuclear Engineering and Design, [Nancy, is this a book or a journal?] 2007. These authors argue that the understanding of predictive capability of a computational model is built on the level of achievement in V&V activities and the quantification of uncertainties related to the application of interest.

the nuclear weapons lab directors for stockpile surety[39] and the Systems Prioritization Method developed as part of the licensing process for the Waste Isolation Pilot Plant[40,41].

The quantification of uncertainty continues as an ongoing research area outside of these communities as well. One of the problems noted by experts in this field is the lack of robust procedures to bridge between theory and practice, with probabilistic methods often being the only framework presented for applied problems. Probabilistic methods, while they have many important advantages, are not be justified in situations of incomplete, heterogeneous, or sparse data, as is often the case for complex social systems[42]. Current research within the Generalized Information Theory community is promising in that it seeks to systematically map existing real world problem situations into the appropriate uncertainty quantification methods via a methodology based on the available data and the requirements of the uncertainty quantification method[43]. This research holds promise for helping communicate to decision makers where uncertainties are in both models and the evidence base for the models, how to reduce the impact of these uncertainties, and how to use the models for improved decision making in light of the inherent and/or irreducible uncertainties.

---

[39] See, for example, written statement of LANL laboratory director John Brown to the Strategic Forces Subcommittee of the Armed Services Committee United States Senate, April 25, 2001. http://www.lanl.gov/orgs/pa/News/testimony042501.pdf

[40] Marietta, M. G., J.C. Helton, N. Hayden Prindle, F. Mendenhall, D. M. Boak, W.E. Beyeler, D.K. Rudeen, R.C. Lincoln, K.M. Trauth, D.R. Anderson, "The Second Iteration of the Systems Prioritization Method: A Systems Prioritization and Decision-aiding tool for the Waste Isolation Pilot Plant. Volume I: Synopsis of Method and Results.", SAND 95-2017/1, Sandia National Laboratories, Albuquerque, NM., May 1996

[41] Prindle, Hayden, N., D.M. Boak, R.F. Weiner, W. Beyeler, S. Hora, M.G. Marietta, J.C. Helton, D. Rudeen, H. Jow, and M. Tierney, "The Second Iteration of the Systems Prioritization Method: A Systems Prioritization and Decision-Aiding Tool for the Waste Isolation Pilot Plant, Volume III, Analysis for Final Programmatic Recommendations", SAND 95-2017/3, Sandia National Laboratories, Albuquerque, NM., May 1996

[42] Sentz, Karl (n/d) "Generalized Information Theory and Practical Applications: Establishing the Methodological Link", Los Alamos National Laboratory, LA-UR 04-4577.

[43] See, for example: George, T. and N.R. Pal, (1996) "Quantification of Conflict in Dempster-Shafer Framework: A New Approach", International Journal of General Systems, Volume 24 No. 4, pp 407-423.; Thow-Yick, L. (1998) "General information theory: some macroscopic dynamics of the human thinking systems", Information Processing and Management, Volume 34, Number 2, March 1998 , pp. 275-290(16); Oberkampf, W.L., J.C. Helton, and K. Sentz, (2001), "Mathematical Representation of Uncertainty", AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Seattle, WA, 16-19 April 2001.; Sentz, K. and S. Ferson. (2002) "Combination of Evidence in Dempster-Shafer Theory." SAND Report 2002-0835, Sandia National Laboratories, Albuquerque, N.M., April 2002.; Klir, G.J. (2003) "Uncertainty-based information." In: Systematic Organization of Information in Fuzzy Systems, edited by P. Melo-Pinto, H. Teodorescu, and T. Fukuda. Amsterdam: IOS Press, pp. 21-52; Agarwal, H., J.E. Renaud, E.L. Preston and D. Padmanabhan, (2004) "Uncertainty Quantification Using Evidence Theory in Multidisciplinary Design Optimization", Reliability Engineering & System Safety, Volume 85, Issues 1-3, July – September 2004, pp. 281-294.; Sentz, K. and A. Wilson, (2005) "Fault Tree Uncertainty Quantification Using Probabilities and Belief Structures on Basic and Non-Basic Events", Fuzzy Information Processing Society, NAFIPS Annual Meeting 26-28 June 2005, pp. 65-68.; Ben-Haim, Y. (2006) "Info-Gap Decision Theory: Decisions Under Severe Uncertainty", Elsevier Publishing Company, NY ,NY.

Summary

The dynamic, co-evolutionary, multi-faceted nature of today's national security threats has encouraged the emergence of many new modeling approaches to support transdisciplinary data collection, analysis, and decision support. In addition, recognition of the need to include the social sciences and the humanities in threat analysis introduces new concepts, frames of reference, and language to what historically has been a field structured around the physical and engineering sciences. The modeling and simulation approaches emerging from this confluence of disciplines hold both promise and challenge, particularly in applications to social systems.

Simulation models of social systems can be useful for numerical computation of analytic models when the solution domain becomes intractable, for testing the robustness of analytic models for sensitivity to certain variables and assumptions, and as stand alone simulations of models of complex, nonlinear systems with no analytic solutions. In this latter case, analytic solutions of parts of the problem may exist, but not for the complex system as a whole where emergent behavior results from interactions among purposeful, self-organizing, actors. Each type of use case dictates a particular approach to verification and validation.

However, irrespective of the particular approach, there are some commonalities to all approaches that should be present. First, the approach should establish upfront the epistemological framework of the analysis, and ensure that the model and simulation methodology selected is appropriate for the analysis purpose, levels of system complexity (including dynamics and co-evolutionary feedback within the system), and data availability. Second, the approach should provide for verification and validation of each step in the overall analysis process (data selection, conceptual model development, computational representation of model, simulation procedures and interpretation of results commensurate with analysis purpose). Uncertainties that are inherent to, or introduced through, each step needs to be identified and the propagation of those uncertainties throughout the entire analysis must be evaluated holistically. Thirdly, the perspectives of users and decision makers must be incorporated in the analysis design with particular attention paid to acceptable uncertainty levels.

Depending on the type of analysis, most validation approaches fall into one of three stages: replicative validity, predictive validity, and structural validity. Of these, it is the latter which is most relevant – and most difficult – for most social simulations. In recent years, the social science simulation community has tackled these issues and has proposed a number of novel approaches for particular applications. Approaches that draw on advances in the fields of artificial intelligence and reasoning, knowledge representation, complex systems, and computational social sciences, have been presented as examples. These examples suggest fruitful areas for exploration, while simultaneously making clear that there is no "one-size-fits-all" method or approach.

Ultimately, the purpose of verification and validation is to contribute to quantification of uncertainty and its sources in the model and use of results. Proven methods of uncertainty quantification developed in the engineering and physical sciences for high risk decision making and in the field of general information theory hold promise for applications to social simulation models.

Verification and validation of social simulation models is a challenging, though not insurmountable problem. Some examples of possible approaches, specific to particular use cases within an overall analytic framework have been presented in this paper. The utility of social simulation models in decision-making situations would be advanced with more in-depth development of a framework that mapped specific modeling methods to validation approaches for classes of use cases and decision-making scenarios.