Priority Challenges for Social and Behavioral Research and Its Modeling

Paul K. Davis, Angela O’Mahony, and Timothy R. Gulden

RAND NDRI/ISDP and NDRI/ATP

WR 1206-DARPA
August 2017
Prepared for DARPA

This document has not been formally reviewed or edited. It may not be cited, quoted, reproduced or transmitted without the permission of the RAND Corporation. RAND’s publications do not necessarily reflect the opinions of its research clients and sponsors. RAND® is a registered trademark.
Preface

This Working Paper was prepared for the 2017 conference of the Computational Social Science Society, to be held in Santa Fe, New Mexico on October 19-22, 2017. The Working Paper draws on a RAND project, the report for which will be published shortly. Informal comments on the Working Paper are welcome and should be addressed to the senior author at pdavis@rand.org.

This research was conducted within the International Security and Defense Policy and Acquisition and Technology Policy Centers of the RAND National Security Research Division (NSRD). NSRD conducts research and analysis on defense and national security topics for the U.S. and allied defense, foreign policy, homeland security, and intelligence communities and foundations and other non-governmental organizations that support defense and national security analysis. For more information on the International Security and Defense Policy Center or the Acquisition and Technology Policy Center, see http://www.rand.org/nsrd/ndri/centers.
Abstract

This paper summarizes priority challenges for social and behavioral modeling based on a recent study building on a base of prior studies. Our focus is less on describing and hand-wringing about the current state than on identifying what is necessary for moving on. Some of the obstacles reflect inherent challenges: social systems are complex adaptive systems; they often pose “wicked problems;” and even the structure of social systems shows emergent behavior. Other obstacles reflect disciplinary norms and practices, mindsets, and numerous very difficult scientific and methodological challenges. We discuss challenges in six groups: (1) tightening links among theory, modeling, and both empirical and computational experimentation; (2) seeking more general and coherent theories while retaining alternative perspectives and narratives, and while effectively confronting multidimensional uncertainty from the outset; (3) assuring that explanatory models represent science faithfully, to include addressing aspects and determinants of behavior that have often been omitted or treated with hard-wired representations; (4) challenging experimenters to find new theory-informed (but not theory-imposing) ways to obtain and analyze relevant data in this modern era of ubiquitous data; (5) challenging theorists and technologists to provide related methods and tools; and (6) nurturing the rest of the ecology needed for overall effectiveness. We suggest identifying several national challenge problems and, for each, having a distributed and virtual social and behavioral laboratory to stimulate synthetic interdisciplinary work. These should feature mixed methods (not just classic simulation) and competition, but also frameworks, modularity, and problem-focused composition—again with competition and evolution—rather than an imagery of standing “correct” federations. Experience shows that breakthroughs often occur as the result of solving concrete problems and then recognizing more general patterns.
1. Introduction

This paper distils the preliminary results of a RAND project seeking to identify priorities for research in social and behavioral modeling and the science that underlies it. The work involves Paul K. Davis, Angela O’Mahony, Rebecca Balebako, Timothy Gulden, Osonde Osoba, and Katharine Sieck. Although the paper is terse, our work built on a foundation of earlier studies (see bibliography), and the full report will have extensive citations to the literature. Feedback on this paper is most welcome.

2. Objectives for Social and Behavioral Research

In the years ahead, social and behavioral modeling should make it possible for researchers and leaders to (1) better understand certain classes of social-behavioral phenomena with national significance; (2) anticipate how those phenomena may plausibly unfold; (3) estimate the potential desirable and undesirable effects of additional events in the world, or of possible U.S. or adversary interventions; and (4) inform pragmatically the work of policymakers and other planners. The phenomena of interest include radicalization for terrorism, weakening of democracy and national cohesion by foreign information operations campaigns, improving prospects for stability after international interventions, managing behaviors of populations after natural disasters, or dealing with opioid or obesity epidemics. The gamut ranges from national security topics to those of a broader character. All of these topics would be good candidates for what we refer to as “National Challenges” later.

Major advances in modeling, however, will be needed if modeling is to contribute as it should in achieving the objectives. We begin by summarizing major shortcomings and obstacles in such modeling—some inherent and some due to current methods and practices. We then identify steps that deserve priority attention.

3. Inherent Difficulties and Challenges

Behavioral modeling in the soft sciences is famously hard. Three reasons merit pondering:

1. Complex Adaptive Systems. Social systems are complex adaptive systems (CAS), which need to be modeled and analyzed accordingly to achieve broader understanding, recognition of patterns and phases, limited forms of prediction, and useful uncertainty-sensitive results (i.e., results shown as a function of context and other assumptions). Great advances are needed in understanding the states of complex adaptive systems, their phase spaces, and recognizing both instabilities and opportunities for influence.

2. Wicked Problems. Many social and behavioral issues arise in the form of so-called “wicked problems,” i.e., problems with no a priori solutions and with stakeholders that do not have stable objective functions. Solutions, if they are found at all, emerge from human interactions.
3. **Structural Dynamics.** The very nature of social systems is often structurally dynamic in that structure changes that may emerge after interactions and events. This implies unusual complications for modeling.

Hard problems, of course, need not be impossible. When they are important but not yet solved, they can be exciting. It is not a pipe dream to imagine valuable behavioral modeling at individual, organizational and societal scales. After all, complex adaptive systems are only chaotic in certain regions of their state spaces. Elsewhere a degree of prediction and influence is possible. It is therefore important to be able to recognize a social system’s state and how subject to influence it is. As for problem wickedness, it should often be possible to understand the generative processes well enough to guide actions (or inactions) that will increase the likelihood of good developments and reduce the likelihood of bad ones.

To be less abstract, consider how experienced leaders can often deal well with crowds that might otherwise become surly (they can also recognize when matters are close to getting out of hand). As for wicked problems, consider how experienced negotiators can sometimes facilitate eventual agreements between nations, or between companies and unions, even when emotions run high and no agreement exists initially about endpoints. Nothing is guaranteed, but experience counts. Analytic methods can also help, as when managers use modeling, gaming, and simulation to anticipate consequences of alternative approaches to reorganization.

So, opportunities exist and—with the benefits of modern social science, technology, modeling, and ubiquitous personal and social data—breakthroughs are possible. They will not come easily.

4 **Obstacles Due to Shortcomings of Current Practice**

4.1 **Fragmented Science**

A natural way to begin improving behavioral modeling is to review obstacles, beginning with the science that should underlie it. Current behavioral theories are many, rich, and informative, but also narrow and fragmented. They do not provide the basis for systemic social and behavioral modeling. More nearly comprehensive and coherent theories are needed, but these will not arise without focused efforts because current disciplinary norms and incentives favor continued narrowness and fragmentation.

Advances will require more interdisciplinary work (including what some authors refer to as trans-disciplinary work), including synthesis at multiple levels of resolution (scales). This is especially demanding because resolution varies along numerous dimensions. As is well known, objects of modeling may be individuals or aggregate groups, but different scales also exist for spatial and temporal matters, the attributes ascribed to the objects, or the interactions among objects. To illustrate, a superficially detailed model might represent millions of individuals, but with behaviors driven by simplistic notions of materialistic rational-actor choices based on unrealistic perceptions and ignoring effects of random exogenous events. In the practice of modeling and analysis, apparent inconsistencies in scale or levels of analysis are often appropriate because not all detail is comparably relevant. Further, model simplifications are essential for many reasons. Thus, the answer is not to go deep in all dimensions, but something more complicated. Understanding when and how to simplify is not yet well understood in social
and behavioral modeling. Fortunately, hints exist about how to gain such understanding from successes in other branches of science.

4.2 Modeling that Does Not Represent the Science Well

Models represent theories and, in language, “theory” and “model” are often used interchangeably. The appropriate type of model will depend on what is being represented and what questions are being asked. Some models may be relatively simple; some may represent the dynamics of complex systems. Some general observations relating to improving the science (not just prediction in a data-rich stable environment) are as follows.

Causality and Uncertainty. To represent social and behavioral theory requires increased emphasis on causal (rather than statistical) models, especially for understanding systems and informing decisions, and on uncertainty-sensitive models that routinely display results parametrically in the multiple dimensions that define context.

Subtle Phenomena. Models need to represent important subtleties of the science, such as those associated with complexity, variable structure, and alternative perspectives (e.g., those of the economist and anthropologist). Some of these are emergent properties, not something to be “baked in.”

Relationships. Although many types and formalisms of models are needed, it is also necessary to know how to relate them to each other—by analogy to how we are able to trace the path between quantum statistical physics and such macroscopic topics as classical thermodynamics and engineering-level theories. It is also necessary to understand relationships between theoretical constructs and both computational and empirical experimentation.

Current modeling falls short on all of these criteria.

4.3 The Need for Modularity and Composition

To move beyond fragmentation, it is necessary to compose complicated models from lower level modules. However, it is not enough to be able to put models together. The composition must be valid for the purpose intended, which can be difficult because the component models may be based on subtly different assumptions.

Meaningful model compositions are necessary for representing real-world systems, which are often nearly hierarchical or modularly networked. The goal, however, should not be monolithic official blessed federations and data bases, but the ability to compose appropriately for a given purpose and context, while representing uncertainty and disagreement, and encouraging competition. Standing standardized federations are likely to be quite problematic.

For such as-needed composition to be feasible, community libraries of well-reviewed modules are needed. Competitive and complementary modules will be needed because science is unsettled and different perspectives are needed. Winnowing is good, but over-standardization is a threat.
4.4 A Myopic View of Model Validity

As noted in past discussions, an obstacle to progress has been an overly narrow view of model validity. Given the nature of social systems as well as modeling limitations, we must rethink “validity.”

- We recommend distinguishing among validity for description, explanation, exploration, coarse prediction, or classic prediction.

Few social and behavioral models will be valid for classic prediction (accurate, precise, and unequivocal), but much higher aspirations are possible for the other purposes. All assessments will depend on context.

5 Strategy for Moving Ahead

5.1 Overview

A way ahead is suggested by Figure 1. The first step should be to define two or three very difficult National Challenge Problems (e.g., based on problems mentioned in Section 2 above) for multi-year efforts forcing good inter-disciplinary work and providing the concrete context that is so important in motivating problem-solving. Unlike grand challenges that pose a single crisp technological feat (e.g., long-distance operation of an autonomous road vehicle), these would have sub challenges in relating to a layered investment strategy. These involve (1) tightening links among theory, modeling, and both empirical and computational experimentation; (2) seeking more general and coherent theories while retaining alternative perspectives and narratives; (3) assuring that models represent theories faithfully, to include addressing aspects and determinants of behavior previously omitted or treated with hard-wired representations; (4) challenging experimenters to find new theory-informed ways to obtain relevant information and analyze data; (5) challenging other theorists and technologists to provide new methods and tools; and (6) improve and nurture the rest of the ecology needed for overall effectiveness. We elaborate in subsequent sections.
For each National Challenge, it may be useful to construct a virtual Social and Behavioral Modeling Laboratory (SBML). An SBML might be seen merely as an organized program akin to the Genome Project of years past, but we use the term laboratory to convey a sense of purposeful and organized scientific inquiry to “crack” a particular challenge. A SBML (Figure 2) would exist for a finite time (e.g., 5-10 years) and would enable interdisciplinary sharing and synergism. An SBML approach would not seek a monolithic standardized model federation with approved structure and data bases. Instead, the approach would be dynamic and iterative with routine competition, iteration, and evolution. Meaningful model compositions would be constructed for specific purposes. The SBML activities would include both simulation modeling (generating system behavior over time) and other forms of qualitative and quantitative modeling, to include participative modeling and such other forms of human interaction as gaming. Related conferences would focus on the National Challenge, the state of related social science, the degree to which M&S represent that science, the products of empirical work and computational experimentation, and how to characterize knowledge and inform decisions. Comparing lessons from multiple national challenges will reveal further generalizations.
Let us now go deeper, beginning with the challenge of improving the theory-experimentation research cycle.

5.2 Tightening the Theory-Modeling-Experimentation Research Cycle

The issue is how to improve interactions among social and behavioral scientists on the one hand, and “modelers” on the other. The problem is implied by the previous sentence, which distinguishes between scientists and modelers. Why are they separate? Physicists, engineers, and economists often do their own modeling. A related challenge is improving the degree to which theories and models can be comprehended, reproduced, peer reviewed, debated, and iterated.

Figure 3 denotes an idealized way of relating the real and model worlds. It comes from a scientific realism perspective but can address many constructivist ideas. The imagery is that a real system exists (the social system of interest, item 1). Real-world observation and experimentation (item 2) help us in abstracting and in forming hypotheses about the elements, relationships, and processes of the system. Because theory and modeling are always simplifications of reality, we must have particular objectives in mind when asking about the real world or how to model it. We may construct one or more rough-cut system theories in our heads to describe the relevant reality (item 3). Often, alternative notions about the system exist, reflecting different hypotheses, perspectives, or both. This is symbolized by the stacking of icons.
Moving rightward, we then construct coherent conceptual models of the system (item 4)—including aspects that are important but not observable directly. The conceptual models may be combinations of essays, listings of objects and attributes, or such devices as influence, stock-and-flow, or network diagrams. The next step, when feasible, is to develop a formal model (item 5), i.e., one that specifies all the information needed for computation of model consequences, even if qualitative. That is, a specified model must have tight definitions, equations, algorithms, or tables.

In this idealized image rooted in classic pedagogy, the formal model is independent of programming language—or more plausible—is expressed in a high-level language that is easily comprehended by non-programmers (i.e., in a week rather than a month). The reason is to lay bare the essence of the model without the mind-muddling complications of most computer code, and to permit replication, peer review, debate, and iteration. After those occur, a formal model can (moving downward) be implemented in other programming languages as convenient (item 6). Moving leftward, the implemented model can then be used with an experimental design to generate model results across the n-dimensional space of model inputs. Results of these exploratory computational experiments (item 7) may falsify or add credence to earlier beliefs and suggest further hypotheses that can enrich system theories. For example, they may suggest that the system will show troublesome behavior in certain circumstances not previously considered. Such coarse computational predictions should be compared to experimental results from the real system (item 8). To do so sensibly requires defining the relevant “experimental frame,” which specifies the conditions under which the real and model systems are to be observed, and how model results are being used. A model can be considered valid for a particular application in a particular context if the result of using the model’s outputs are adequately close to the result of
using the real system’s behavior for that experimental frame. The cycle continues (dotted arrow). Overall, Figure 3 is a virtuous research cycle.

Although analogous to depictions in influential academic literature on system modeling, Figure 3 has special features. First, it is oriented more to science than engineering: it allows for research objectives that are fuzzy, as in “understand what is going on.” Second, it anticipates diverse types of models, sometimes reflecting fundamental differences of perspective (e.g., those of an economist and an anthropologist or sociologist). Third, it shows pivot points at which theory affects empirical and computational experimentation, and comparisons between theory and experiment. The theories inform what experiments should be conducted and how results should be analyzed. Finally, the SBML concept of Figure 3 takes seriously the separate existence of conceptual and specified models. This is analogous to representing a model in mathematics before programming, but recognizes that some models are not reasonably expressed in mathematics and would be incomprehensible if they were. The need for this step is a matter of dispute and alternatives should indeed be defined and debated.

Figure 3, then, encourages a holistic view in which theory, models, empirical inquiry, and computational inquiry are all part of a closely networked system. What follows describes suggested priorities for improving other aspects of the science, the methods and tools that enable improvements to science, and other aspects of the ecosystem.

5.3 Improving Theory and Related Modeling

Some priorities for social and behavioral research appear to be as follows, expressed as objectives at the functional level.

1. Move modeling farther into the science itself. An example might be an anthropologist professor and graduate student doing the modeling directly or accomplishing this with exceptionally tight teaming across departments. All concerned would be intimately familiar with the conceptual and formal models. This contrasts with a modeler first building a model and only then, annoyingly, calling upon social scientists to estimate input values.

2. Put a premium on causal models that include all important variables, whether or not “soft.” Assure that models are faithful to the emerging theory, including allowing system characteristics to be emergent. As part of this, develop methods for variable-structure simulation, to include versions in which changes are emergent.

3. Define context. Go beyond hand-waving reference to “it all depends on context” by defining what characterizes context—e.g., state variables, exogenous events, and aspects of history.

4. Build in uncertainty analysis from the outset. Build uncertainties about model structure as well as model-parameter values into the very fabric of models. Prepare for analysis under deep uncertainty.

5. Use portfolios of analytic tools. The portfolio should include simulation, but also, e.g., equilibrium models, static models applicable at a point in time, knowledge-based simulation, qualitative models, and such human-in-the-loop mechanisms as games.
6. *Seek families of models* that cut across levels of detail and perspective (multiresolution, multiperspective models). Examples include individual versus population levels, or immediate versus long time scales. Part of this will involve respecting different views of phenomena, including culture-sensitive views.

7. *Synthesize across theories* where feasible. A more nearly general theory may say, for example, that a behavior can be due to any one or a combination of pathways dependent on numerous variables. This may integrate pathways previously treated as separate theories.

8. *Translations.* Sharpen understanding of where useful theories are equivalent but expressed in different languages or formalisms, and where they are inherently incommensurate (e.g., individual-centric and culture-centric models).


10. *Organize for interdisciplinary work,* both collaborative and competitive.

### 5.4 Improving the Computational and Empirical Data

A somewhat shorter list of admonitions applies to improving the data obtained from real-world experiments and computational experimentation.

1. *Seek the right data even if inconvenient.* Assure that empirical data is appropriate for purpose (i.e., the National Challenge). In particular, address difficult-to-measure and sometimes latent qualitative variables, seeking representative albeit uncertain data rather than poor proxies while dealing with data biases, data requirements for alternative theories, and uncertainties.

2. *Exploit opportunities.* Better exploit modern and emerging data sources, which are sometimes massive, sometimes sparse, and often riddled with subtle correlations and biases?

3. *Shorten the cycle.* Greatly increase the speed with which needed data can be obtained and processed.

4. *Giving theory a chance.* Use theory-informed approaches wherever possible when designing and conducting empirical or computational experiments. These should not be theory-imposing; but they should stress testing and enriching theories. That said, healthy competition should exist between, data-driven and theory-informed work.

5. *Use exploratory analysis.* Analyze computational data with the methods of exploratory analysis, looking for phase diagrams rather than point outcomes.

### 5.5 Improving Methods and Technology for Modeling and Experimentation

The priorities expressed above assume methods and technology for modeling, data collection, and analysis. Some of what is needed does not yet exist. Some priority questions for providers of modeling theory, methods, and tools are how do we

—Build the capacity for multidimensional deep-uncertainty analysis into the very fabric of the analytic tools?
—Build variable-structure simulations in which agent types, the character of their reasoning, and their parameters change or even emerge within the course of simulation?
—Identify and define the suite of methods and tools are needed to represent and compare different scales and alternative models with different narratives or perspectives? How do they relate?
—Conduct theory-informed empirical and computational experiments? How do we design, collect, and analyze?
—Make the best use of modern data sources, which are sometimes massive, sometimes sparse, and often riddled with subtle correlations?
—Infer likely causal relations from only partially controlled observational data?
—Accomplish heterogeneous fusion and otherwise “manage knowledge” so that information can be used despite variations of scale, formalism, and character?
—Build conceptual and formal models that are comprehensible to subject-area scientists and subject to reproducibility, peer review, debate, sharing, and iteration?
—Reconceive “composability” to include accommodating analysis under multidimensional deep uncertainty, and the search for robust conclusions and strategies? How do we improve prospects for composability by improving comprehensibility and the ability systematically to check for composition validity?

Another set of questions is what output should look like along the way and when informing decision makers or the public? How can communication be improved when dealing with multi-dimensional issues while balancing depth and breadth, and using cognitively effective mechanisms?

5.6 Attending to Culture, Governance, and Other Ecosystem Issues

Although the items that follow are less about research than about management issue for government, the overall effort needs to address all aspects of the challenge’s ecostructure.

Cultural Mindsets. Mismatches currently exist—among both recipients of research and the researchers themselves—between what is sought and what should be sought in describing knowledge well and informing reasoned decision-making. Mindset changes are needed, especially in the domain of social and behavioral issues:
—From seeking narrow point predictions and optimization (akin to engineering) toward seeking analysis of wicked problems and development of robust strategies.
—From seeking simple once-and-for-all strategy toward explicitly adaptive strategies that require routine monitoring and adaptation.
—From seeing “human terrain” as something to characterize with solid, fixed date toward seeing it as complex, heterogeneous, and dynamic—in structure, not just data values.

These changes would be consistent with modern research in the decision sciences.

Infrastructure. The relevant elements of infrastructure include vigorous academic and private-world research programs in the US and other countries, related funding, and mechanisms for interaction. It includes laws, regulations, funding strategies, and relationships that encourage balances among, e.g., openness, sharing, and protection of intellectual property and privacy rights (see below).
Governance and Ethics, Privacy, and Civil Liberties. Another class of issues involves the ethics of assuring privacy and respect for civil liberties. Many elements of research will involve data collection and data analysis, some of which could cause difficulties and, in some instances violate the intentions of law. It is critical to maintain and help establish the highest standards on such matters, but to do so will require continued analysis and innovation, and perhaps promulgation of suitable laws.

Acknowledgment and Disclaimer

This paper stems from a research project funded by the Defense Advanced Research Projects Agency. The views, opinions, and/or findings expressed are those of the author(s) and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

Bibliography

Egeth, Jill D., Gary L. Klein, and Dylan Schmorrow, (eds.) (2014), Sociocultural Behavior Sensemaking: State of the Art in Understanding the Operational Environment, McLean, Virg.: MITRE.


We were also influenced by past RAND work with which we were particularly familiar dealing with multiresolution modeling, model composability, exploratory analysis, causal social-science modeling, and transitions from qualitative to semi-quantitative modeling and analysis. See, e.g, Davis, Paul K. and Angela O’Mahony (2017), “Representing Qualitative Social Science in Computational Models to Aid Reasoning Under Uncertainty: National Security Examples,” Journal of Defense Modeling and Simulation, 14(1), 1-22.