An Artificial Intelligence/Machine Learning Perspective on Social Simulation

New Data and New Challenges

Osonde A. Osoba and Paul K. Davis

RAND NDRI/ISDP and ATP

WR1213-DARPA
December 2017
Prepared for DARPA

Distribution Statement A (Approved for public release, Distribution Unlimited)

RAND working papers are intended to share researchers' latest findings and to solicit informal peer review. They have been approved for circulation by RAND National Defense Research Institute but have not been formally edited or peer reviewed. Unless otherwise indicated, working papers can be quoted and cited without permission of the author, provided the source is clearly referred to as a working paper. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. RAND® is a registered trademark.

Distribution Statement "A" (Approved for public release, Distribution Unlimited)
Preface

This Working Paper was prepared for an edited volume on social behavioral science modeling to be published in early 2018. 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 Working Paper reviews the current state of the art in data infrastructure and artificial intelligence approaches that could be valuable for social and behavioral modeling. Among the newer machine-learning methods, adversarial training and fuzzy cognitive maps seem to have particular unrealized potential.

The Working Paper then discusses the troublesome theory-data gap: the mismatch between measurable data streams and meaningful explanatory theories to frame the data. The Working Paper identifies this issue as a key barrier to meaningful social and behavioral modeling. It then discusses the need to move from purely data-driven work to theory-informed work, and to tighten the iterative loop between theory and data analysis.

Closing the theory-data gap is a general problem. The Working Paper illustrates three example models that attempt to integrate theory-informed and data-driven modeling: a network model, factor tree model, and a fuzzy cognitive map model. The first model addresses meme transmission. The last two address the public support for terrorism.

The Working Paper identifies key questions and challenges along the way. These are questions that a notional social and behavioral modeling research community will need to tackle as it grows.
1 Introduction

1.1 Objectives

There is a growing demand to develop social and behavioral models competent to inform decisionmaking in such diverse domains as counter-insurgency, political polarization, adversary propaganda campaigns, and public health behaviors. Success will depend on effective use of empirical information drawn from observation of both online- and physical-network human behavior. The objectives of this discussion are

1. to characterize the current infrastructure of data relevant to behavioral modeling;
2. to describe progress on methods relevant to behavioral modeling that come from research on artificial intelligence (AI) and machine learning (ML);
3. to identify shortcomings and challenges with current modeling approaches;
4. to suggest where advances are needed to address them;
5. to identify some mechanisms for doing so.

1.2 Background

An earlier study conducted over three years by the National Research Council (Zacharias et al., 2008) presented a comprehensive review of social and behavioral modeling. Most of that analysis remains solid and apt today. Thus, we focus primarily on selected developments over the last 10 years.

2 Relevant Advances

2.1 Overview

Two trends relevant to this Working Paper have been notable over the last decade (1) the burgeoning of data, data sources, and data infrastructure; and (2) advances in intelligence/machine learning (AI/ML) methods.

Data: Social Media More Than MMOGs. The 2008 NRC report highlighted Massive Multiplayer Online Games (MMOG) platforms, surveys, and ethnography as predominant sources of behavioral data for training behavioral models. Some recent work has explored MMOGs for understanding observed social behaviors (e.g., identifying patterns observable in the play of Pokémon Go (Althoff et al., 2016; Althoff et al., 2017)). More generally, however, it seems that the usefulness of MMOG platforms for behavioral insight is less than was imagined.
MMOGs are interesting troves of data, but the insights appear not to add much to the foundations of behavioral modeling research. This observation makes sense in hindsight. MMOGs are specific ecosystems of behaviors, and contrived ones at that. Researchers can distil out interesting patterns of behavior (Tomai et al., 2013), but these are not necessarily generalizable or informative outside the contrived environment. Arguably, surveys and ethnography (including micro-narratives) are better for eliciting insights on more general behavioral patterns, although they can be unwieldy and intrusive. Ubiquitous social media platforms seem to be popular as sources of relevant behavioral data for now. We shall discuss some of this in a later section.

**AI/ML Methods.** As for advances in AI/ML methods, we construe the topic broadly to include:

- Adaptive statistics- and optimization-based methods for teaching computers to identify or exploit regularities in signals (Machine Learning or Pattern Recognition)
- Expert-, Rule- or logic-based methods for planning, problem solving or knowledge representation.
- Models for representing or imitating cognition and decision-making (human or otherwise)

These topics fall under the useful general definition of AI as the discipline “concerned with intelligent behavior in artifacts”.

The term “machine learning” (ML) often refers to the more statistically flavored sub-fields like supervised, unsupervised, & reinforcement learning. Other AI sub-fields rely more on symbolic-based and rule-based methods for tasks like knowledge representation and automated planning. Some earlier versions of these were known as expert systems and knowledge-based systems in the 1980s. These include AI approaches like automated planning solvers, fuzzy cognitive maps (Amirkhani et al., 2017), and tree-based methods for parsing semantics and ontologies. Other non-ML strands of AI research include cognitive modeling architectures like BDI (Tambe et al., 1988), Soar (Georgeff et al., 1998), ACT-R (Anderson, 1996), and EPIC (Rubinstein et al., 2001). These have been useful for enabling tasks like team-based collaboration in robots. The varied nature of social and behavioral modeling requires the full diversity AI methods. We see it as important to consider all of these methods (i.e., to include what some refer to as both strong and weak strands of AI).

---

1 There have been numerous other attempts to define AI canonically. This description is due to Nilsson (Nilsson, 1998). McCarthy (2007) defines intelligence as “the computational part of the ability to achieve goals in the world.” Minsky (1961) gave an enumeration of functions required for such intelligence: search, pattern recognition, learning, planning, and induction (or generalization from observed examples). Any artificially constructed or software-based system performing combinations of these functions to achieve goals in the world will qualify as AI for the purpose of our discussion.

2 Strong AI aspires to computer programs that represent human cognition and achieve significant aspects of human-like intelligence. AI researchers are strongly divided about the degree to which strong AI is feasible.
Numerous relationships exist among what are sometimes treated as different methods. The authors of the 2008 NRC report generated one depiction of the various methods and how they relate to each other, as indicated in Figure 1. Although useful for drafting and structuring the large and complex NRC report, Figure 1 uses a fine-grained disaggregation of modeling approaches, which the report discussed as Individual, Organizational, and Societal (IOS) modeling tools. For our purposes, such a disaggregation downplays how deeply interconnected these approaches sometimes are and sometimes should be. For example, the optimization node (near bottom left) stands alone in the figure. However, a central theme in current ML methods is learning-as-optimization (see dashed line 1); much of modern ML relies on optimization procedures like stochastic gradient descent during training. The distinction between machine learning and statistics is also not nearly wide as the figure suggests (see dotted line 2). Another problem with such a disaggregation is that it may obscure opportunities for innovation. An important recent innovation in AI/ML, generative adversarial networks (GANs), combines game theory and machine learning to improve unsupervised learning tasks. Other current innovations include the use of statistical, ML, and social network analysis methods to infer behavioral patterns (Sapiezynski et al., 2016).

![Figure 1 A Similarity Network of Modeling Methods](image)

Source: Adapted from NRC report (Zacharias et al., 2008) p. 93

The subsequent sections discuss advances in data sources and AI/ML in more detail.
2.2 Advances in Data Infrastructure

2.2.1 New Sources

Many new data sources have emerged in the last decade as important ecosystems for exhibiting and recording behavioral patterns. Each data source has blind spots. On the one hand, a larger ecosystem might be expected to increase the chances of capturing key behavioral information. On the other hand, a smaller data ecosystem would require less modeling effort and might be more cost efficient because infrastructure costs burgeon with ecosystem size, perhaps faster than any benefits. Nonetheless, from a purely modeling perspective, one might think that more data from more sources should—other things being equal—increase capacity for behavioral modeling.

Some novel data streams have already shown value for behavioral modeling. For example, social media (SM) data has become ubiquitous and is now an important analytic tool in national politics. Social-media platforms also serve as tools for influencing discourse and for measuring behaviors/influence. Social media platforms measure individual data and also relational or network data. The ascendance of social media platforms has led to a rise in methods and tools for dealing with relational/network/graph information.

Cellular Data Records (CDRs), including metadata are also valuable, especially for inferring spatial behavioral preferences. Recent research indicates that CDRs are very useful for identifying spatial behaviors relevant to, e.g., migration/disaster response (Bengtsson et al., 2011), shopping patterns (De Montjoye et al., 2015), and personal network affinities. The recent-year developments related to “WikiLeaks” demonstrate the real-life significance of even limited metadata from cellular data records.

Other novel data streams include:
- financial records from banks, retail records, and financial technology firms (FinTech)
- public and private video surveillance from street cameras, and cellphone recordings
- voice records: from AI personal assistants (Siri, Alexa, Google Now)

The growth in behaviorally relevant data streams mirrors a growing reliance on tools and devices for mediating behaviors (e.g. GPS for navigation, music streams for mood management, social media platforms for expression). It is now common to have records or signals of an individual’s plans, intentions, and mental states in digital form, especially in affluent cultures with high smartphone adoption. Clark and Chalmers (Clark & Chalmers, 1998) used the term “The Extended Mind” to refer to the extension of mental deliberation and cognition outside the (as yet) unobservable confines of the human mind. Extended minds with accessible digital data exhausts can be potentially revolutionary for behavioral modeling. The data ecosystem seems to be growing in that direction.
2.2.2 Evaluating the Data Ecosystem

Whatever its primary function may be, for our purposes a data ecosystem serves a measurement function. It value in this role depends on at least the following:

1. **Representativeness of the measured population**: the measured population (the proxy population) should be representative of the background population about which we want to infer behavioral patterns. Unfortunately, social media platforms generally exhibit significant *proxy population mismatch* (Ruths and Pfeffer, 2014). Proxy population mismatch has been implicated as the primary reason for the errors observed in social media polls (Chung and Mustafaraj, 2011; O’Connor et al., 2010; Gayo-Avello, 2013).

2. **Signal Fidelity & Resolution**: The value of a measurement tool depends on its ability to indicate signals of interest faithfully. Can the data unambiguously indicate behavioral signals of interest? At reasonable effort? Linguistic interactions (e.g. on social media platforms), for example, may not recognize behavioral signals such as sarcasm without considerable effort to parse the data. Also, the ability to usefully indicate location or geographical behavior depends on the resolution of the location sensors. Signal fidelity also includes questions of *misrepresentation* in observed signals. The ability of agents (human or organizational) to harbor unrevealed or unstable preferences and engage in game-theoretic behaviors means that dishonest signaling can be prevalent in behavioral contexts. Behavioral models will need to account for misrepresentation in signals.

3. **Systemic Selective Non-response**: Silence—i.e., the absence of data—is sometimes extremely informative, especially for active voluntary interactions. Unfortunately, the tendency in analysis is to emphasize signal presence over absence. This tendency has been implicated as a potential cause of excessive polarization in social media interactions. Studies of political discourse on blogs and social media platforms highlight the tendency for hyperpolarized minorities to commandeer discourse and thereby skew platform data streams away from underlying population preferences (Tumasjan et al., 2014).

2.2.3 Trends

The data ecosystem is likely to keep growing, perhaps even exponentially as the Internet of Things (IoT) takes off. The goal is not just to obtain larger quantities of data, but also to obtain more raw/unfiltered qualitative & quantitative data, free of the self-report and interpretive biases that often afflict surveys and ethnographies. This ecosystem skews heavily towards observational
Data as opposed to controlled trial data that would identify causal connections more readily. This limits the fitness of the data ecosystem for some purposes.

Questions of data quality, fitness for use, and access will be key. Much of the data ecosystem exists under different jurisdictions, access/legal restrictions, and quality levels. There will likely be a growing demand for methods and practices for fusing data sources and patching blind spots in the ecosystem. Advances in AI/ML, discussed below, may help. Besides those discussed below, the research community is also developing important tools like quasi-experimental approaches for inferring causality from observational data and fusion methods for fusing heterogeneous data/knowledge/information sources. These could be invaluable in time.

Although we do not dwell on the matter here, substantial questions and concerns exist about the rules and regulations governing the use of such data e.g. questions of data privacy, data representativeness, and legal access to data (Levendowski, 2017).

2.3 Advances in AI/ML

Many of the advances in AI/ML have occurred in AI sub-fields that use statistical learning concepts (pattern recognition including advanced regression, clustering, classification). Expert systems, rule-based, and solver-based AIs continue to be important for problems like planning and knowledge representation, but progress there has not gotten as much popular attention. We mention them nonetheless. For example, we present an expert system, a fuzzy cognitive map (FCM) (Osoba and Kosko, 2017), for modeling behaviors later in this Working Paper. Fuzzy-based expert systems also have robust mechanisms for knowledge fusion. Such fusion approaches hold promise for addressing the issue of limited model federation (which we highlight later). Other AI sub-fields like knowledge representation may be useful for specifying underlying behavioral ontologies.

2.3.1 Deep Learning

Deep learning (DL) is the most touted recent trend in machine learning. Traditional connectionist machine learning models, by definition, connect nonlinear processing (neural) units configured in shallow hierarchies or layers to solve classification, regression, or dimension reduction tasks. Shallow networks are limited in the complexity of features (or combinations of input variables) they can find and use. But deeper hierarchies are harder to train. Deep learning leverages advances in computational power and statistical learning theory to update the standard connectionist learning models with deeply stacked layers of processing units. The use of deep stacks allows the model to identify complex features in the data that can be useful for improving the model’s performance.

Depth in learning architectures is an idea that has been considered for much of the history of machine learning research. The main hurdle preventing the exploitation of deep architectures has been limits in usable computing power. Machine learning models solve an optimization problem
in the process of learning. The optimization problem is a function of the number of tunable parameters in the model. Neural network models (both shallow and deep) typically rely on the backpropagation algorithm for parameter tuning. Shallow models have fewer parameters than deep models. The difficulty of the optimization task grows exponentially with the number of parameters. So deep models can be prohibitively computationally intensive. Deep models often also require larger data sets for training. The key factors that make DL models feasible are the existence of large application-relevant data sets and massive computing power. Our previous discussion already highlighted the growth of the ecosystem of behavior-related data. Available computational power has also grown explosively.

More recent DL advances include the use of time varying layer weights that allows models to incorporate temporal memory. These are called recurrent neural nets (RNNs). These models are useful for modeling time-series with temporal correlations. Depth in these models refers to the length of time-dependence in the signal, not number of layers. Long short-term memory (LSTMs) models are a popular type of RNN model. LSTMs are used for handwriting recognition in some Windows systems. They have also shown good results in the generation of text for dialog (e.g. chatbots). LSTMs are especially useful for sequence-to-sequence learning tasks (Sutskever et al., 2014) (e.g. machine language translation like the Google Neural Machine Translation (Wu et al., 2016)). The DL community has also made widespread use of the convolutional neural network (CNN) architecture for video, image, speech, and text tasks. CNNs are neural networks with layers that apply biologically inspired weighted local averages (or convolutions) to input signal fields. CNNs are especially effective at image tasks.

The value of DL for social and behavioral modeling lays mainly the ability of DL models to convert a larger portion of the data ecosystem into behaviorally meaningful signals. The semantics of images and videos used in social interactions become more accessible using the appropriate DL model. The next section talks about natural language processing (NLP). Much of the recent advances in NLP are due to the application of DL models to language tasks. DL has also fostered the development of computing architecture for scalable computation on large data sets. But DL models themselves have not had a large footprint so far as tools for modeling social behavior directly.

2.3.2 Natural Language Processing (NLP)

The most important AI improvement relevant to behavioral modeling over the last half-decade has probably been the maturation of natural language processing (NLP). This includes text mining, topic modeling, sentiment inference, speech recognition, machine translation, etc. These are important because they enable the quantitative analysis of textual data. Behavioral cues in language use are now observable. Caliskan-Islam et al. (Caliskan-Islam et al., 2015; Caliskan-Islam et al., 2016) demonstrates an example of behavioral modeling (identifying telltale patterns of language use) based on natural language processing.
Topic modeling (Blei, 2012; Blei and Lafferty, 2006) has been particularly useful for measuring trends in what is otherwise unstructured data. Blei gives an example of the use of NLP methods to track publishing behaviors in academic fields (Blei, 2012). Topic modeling tools have also been indispensable in the analysis of micro text corpora from social media platforms (e.g. short tweets). Some social-media studies use Latent Dirichlet Allocation (LDA) for natural language processing (Gross and Murthy, 2014).

Other modes of text analysis are needed. Topic modeling is a strictly statistical analysis of text with limited use of semantics or lexical structure. It treats text as pure symbols with no meaning separate from the statistical co-occurrence patterns learned from corpora. Significant insight can be gained from this purely symbolic analysis (as its current use in text mining indicates), but topic-modeling results still require significant human interpretation to identify understand context-dependent meanings. Automated tools for semantic, lexical, and lexico-grammatical analysis of larger bodies of discourse can improve social and behavioral SB models. The rise of NLP suites of tools like word2vec and GloVe represent increasing capacity for the algorithmic comprehension of text meaning. They are currently good at solving analogical questions – arguably the minimal task required to indicate semantic comprehension.

Automated Machine Translation (AMT) is another NLP domain with significant innovation. Google recently introduced the Google Neural Machine Translation (GNMT) (Wu et al., 2016) for language translation using an end-to-end neural network framework trained in a purely data-driven fashion (no pre-coded language rules) (Sutskever et al., 2014).

2.3.3 Adversarial Training for Unsupervised Learning

The discovery of Adversarial Training approaches may hold the most promise for future social-behavioral modeling and related AI. Goodfellow et al. first introduced adversarial training to the AI community in their development of Generative Adversarial Networks (GANs) (Goodfellow et al., 2014; Radford et al., 2015). The training approach takes inspiration from game theory and minimax decision-making.

The standard statistical learning approach begins with training data and a parameterized model (e.g. regressions, perceptrons, support vector machines, or neural networks). The goal is to learn a desired behavior encoded in the training data. Changing the model parameters changes the behavior of the model. Learning algorithms encode the behavior to be learned as critical points of an objective function over the model’s parameter space. Statistical learning (clustering, fitting, classification, regression being the major forms) thus often reduces to stochastic optimization (Vapnik, 2013). This mode of automated learning works well for problems in which the desired behavior is reasonably articulable as objective functions e.g. image/facial recognition and some aspects of natural language processing. This is essentially a teacher-student learning model.

Learning to generate more subtle but sometimes crucial behaviors from data sets can be much harder. For example, generating believable super-resolved images, social networks,
conversations, or even facial expressions is difficult. The key theme in these tasks is the need to learn implicit models of data generation (Mohamed and Lakshminarayanan, 2016; Diggle and Gratton, 1984). These are models for which the generative structure is apparent but not easily articulable. And thus the process of learning the structure consists of model adaptation in response to a repeated game of discriminating between examples.

Adversarial training proceeds as a game between two agents: the generator and the discriminator. The generator’s goal is to learn to produce samples that are representative of the training data. The discriminator’s goal is to learn to discriminate between the generator’s output and samples from the training data. Both agents escalate in the course of training to the point where the generator has learned to behave indistinguishably from the training data. The learning model is essentially an actor-critic learning model (Pfau and Vinyals, 2016). Adversarial training provides a way for AI systems to do better in games-against-nature scenarios or in adversarial scenarios (e.g., individuals are seeking to hide information or even mislead). Such training could enable agents in social science models to learn behaviors that are not easy to encode or explain (e.g., identifying fake news or recognizing patterns in the presence of efforts to hide them or deceive).

2.3.4 Reinforcement learning

Reinforcement learning (RL) is a branch of statistical machine learning focused on teaching agents how to act to achieve goals in an uncontrolled environment (Sutton and Barto, 1998). It has its origins in research on control theory, robotics, automated planning, and behavioral psychology. RL’s key defining features are the explicit modeling of the environment, the built-in emphasis on exploration, the sparsity of evaluative feedback to guide the agent’s learning, and the learning of action policies from experience or data. The rise of RL is a response to the inadequacies of supervised learning for planning-style tasks in which the value of real-time actions derive from their downstream effects rather than immediate evaluations. Planning is an integral part of human social behavior. And we need models that can capture such behaviors.

Figure 2 Standard Reinforcement Learning Framework (adapted from (Sutton and Barto, 1998))

The maturation of reinforcement learning enables the development of large-scale agent based models (ABMs) with adaptive or learning agents (including agents that learn or adapt based on...
simulations). Developments in reinforcement learning may allow for more adaptive ABM models of intelligent human behavior in complex adaptive systems.

2.3.4 Emulating Human Biases and Bounded Rationality

Another important area of work on AI/ML deals with methods for learning true or revealed preferences. This is important for behavioral modeling because self-report data and behaviorally revealed preferences often diverge (Rudder, 2014). Biases in preferences drive many social behaviors. Research on cognitive biases, as reviewed in (Kahneman, 2011), shows that biases perform useful functions even if they sometimes lead users astray. More specifically, recent work by Pita et al. (Pita et al., 2010) on security games shows the importance of accurately modeling human bounded rationality (Simon, 1996; Simon and Newell, 1962). Eliminating heuristic biases (or even irrational aspects of decision-making) from social-behavioral simulation models may reduce their descriptive accuracy. The social simulations need to reflect true biases and preferences.

Researchers are beginning to develop simple games and other methods to discover hidden preferences. These may help address the NRC concern about reproducing non-rational agent behaviors in models (Zacharias et al., 2008, p.359). It is also important, as discussed in the literatures on deterrence (National Research Council, 2014, p.37) and “wicked problems” (Rittel and Noble, 1989) to recognize that humans (all of us) often do not actually have the stable utility functions postulated by both rational-actor theory and usual versions of bounded-rationality theory. We often discover or develop values in the processes of human interaction and experience. “Limited rationality” is sometimes used to refer to problems that go beyond those of bounded rationality (Davis, 2014b, p.6). Such problems include mental-health problems (as when leaders are perhaps depressed and using alcohol or drugs, and are sometimes paranoid) and emotion-driven irrational acts (op cit).

2.3.5 Trends

Most of the AI/ML methods highlighted seem to have more promise as tools for comprehending novel data streams than as new approaches to social and behavioral modeling. The application of RL to ABMs may prove to be the exception in time. But the addition of heterogeneous data comprehension and fusion capabilities to the behavioral modeling repertoire is substantial boon, if nothing else.

Perhaps a key limitation of many currently popular AI/ML methods is the over-representation of feed-forward model structures. This limits the use of such methods to sub-tasks of the behavioral modeling enterprise (e.g. NLP). Training statistical models that feature feedbacks can be daunting or unstable. But many interesting behaviors involve significant amounts of feedback. The feed-forward emphasis also inhibits the representation of chaotic dynamics that can often occur in behavioral models (e.g. in simple predator-prey models (Schaffer 1985)), especially complex ones. The introduction of ML models like GANs and
LSTMs marks a growing emphasis on simple feedback in ML models. Some less popular AI models like FCMs for modeling causal networks also model feedback since causation often features feedback loops.

3 Data and Theory for Behavioral Modeling and Simulation

3.1 Prefacing Comments on Fundamentals

Certain foundational questions must be addressed if we are to model and simulate social behavior. These include how to represent aspects of behavior, how to understand and establish “validity,” and how to compose models from smaller models (model federation). We do not address all such questions here. Here, we focus on the question of: how to relate theory and data.

The relationship between theory and data demands further scrutiny because of the advent of “big data”. Advances in AI/ML, driven in part by the glut of big-data sources, has pre-disposed modelers to focus on data as the primary foundation for building and validating models of behavior. This data focus is evident in the emphasis on predictive power in much of modern machine learning approaches to modeling. The growth of ML modeling approaches is arguably due to the ability to convert measures of predictive validity into concrete mathematical objective functions (e.g. classification accuracy for classification, mean squared error for regressions) that are directly amenable to optimization methods. Explanatory power, if considered at all, has been a secondary goal—sometimes with the rationale that increases in predictive power “surely” occur due to increases in the intrinsic or effective explanatory power of the model, even if that increased explanatory power is difficult to verify because the model complicated or opaque. That rationalization is neither valid nor acceptable. Explanatory power is extremely important for social and behavioral modeling, especially when that is to be used to inform decision-making (Davis et al., 2017). A purely data-driven modeling approach would be inadequate and potentially misguided.

3.2 For want of good theory…

Researchers have multiple examples highlighting the inadequacies of purely data-driven models in complex domains. Recent discussion (Jonas and Kording, 2017; The Economist, 2017) argues that many data-driven neuroscience techniques aimed at uncovering brain architecture and function are misguided. One example involves the common approach of localizing function in the brain based on observations of function loss or impairment due to localized brain lesions. The authors apply modern data-driven neuroscience techniques to an older simpler microprocessor. The goal was to reconstruct the known architecture of the microprocessor by analysis of measured signals and localized impairments or interventions— the way a
neuroscientist might seek to reconstruct the architecture of brain. The methods failed to identify high-level structures (e.g. the arithmetic and logic unit) fundamental to explaining the microprocessor’s function. The failure of the methods at identifying crucial structures in a simpler system raises questions about the ability of these methods to make sense or explain brain functions physiologically.

An older critique of heavily data-driven models of behavior is Chomsky’s critique (Chomsky, 1959; Fodor, 1965) of B.F. Skinner’s model of learned verbal behavior. Skinner had put forward a theoretical model of language acquisition in which infants learn language entirely from experience filtered through a sparse operant conditioning learning framework. Chomsky argued that Skinner’s theory of verbal behavior was wrong because the stimulus-response mechanism it hinged on was too sparse to account for the speed of language acquisition and the observed complexities of language use. He further argued that the human facility with language likely exists because of an innate grammar acquisition model more complex than operant conditioning.

Efforts at building social and behavioral models are subject to similar misspecification risks. Data-driven behavioral models, without the benefit of strong hints from plausible social-science theories of behavior may have poor explanatory power—especially in nonlinear systems. The statistical models may not even be useful for postdiction or post-hoc sense-making. To be sure, fully theory-driven modeling can also be seriously misguided.

In an ideal situation with complete relevant theory and perfect and complete data, the data validates the theories and the theories provide an explanatory/interpretive frame for the data. The current situation, however, is very different: we have a lot of data but it is often imperfect, incomplete, and/or biased, and we have a great many social-science “theories,” which are often narrow, fragmentary, unvalidated, and certainly not settled.4

Key Questions: How do we design models that are valid for purpose and representative of the relevant reality when the data and theories are unsettled? What is the right balance of theory and data focus in models? And what are the best practices on weaving the two together effectively?

3 Mathematically, a core problem is that statistical methods do not “discover” model fragments that are not part of the specification used.

4 It is notable that in the social sciences, “theory” may correspond to nothing more than a single-variable hypothesis, such as “More of X should tend to increase Y,” whereas in the physical sciences “theory” often (but not always) is regarded as that which pulls together a great many considerations coherently. Thus, in conversation, a physical scientist may use the word “theory” in a very positive way, whereas a social scientist may regard someone else’s “theory” skeptically, as nothing more than yet another hypothesis to be tested. Miscommunication occurs routinely on this matter.
3.3 The Scope of Theory and Laws for Behavioral Models

Some earlier literature on the validity of modeling and simulation (M&S) often supposed that complete systems understanding is the key to good models. In other words, good models should have structural validity. This is not necessarily true; good theories are necessary but they often do not need to be perfect, complete, or structurally isomorphic to the real system. This point will be important as we attempt to tease out the laws and theories needed for good social-behavioral modeling and simulation.

The most persuasive way to argue this is perhaps to appeal to daily experience. We do not use detailed mental models of the world to function successfully. The taxi driver does not need a comprehensive mental model of car mechanics or of petro-chemistry to operate his taxi. Similarly, animals do not need a full understanding of physical theories (e.g. Newton’s laws or quantum mechanics) to thrive.

So also, an agent’s internal models of and behaviors in complex systems need not be perfect. They should, however, reflect effective theories (Randall, 2017) to guide their behavior in the system over time. These are simplified mental models that have proven sufficiently useful for guiding the agent’s actions even though they are not fully accurate or complete.

**Key Questions:** Is the social science research community able to identify which theories of social behavior are effective theories vs. accurate theories vs. just interesting-but-invalid theories? Is there a stable mechanism or framework for making this distinction? How useful is data for this purpose?

This value of effective theories for useful modeling suggests a diagnostic concept for modeling practice: “the model-inquiry gap.” This is the conceptual gap between the level of a model’s representation of the world and the level of analysis the model is meant to inform. It is one aspect of estimating how well a model matches its purpose. One example of a large model-inquiry gap would be using a quantum mechanical model in an attempt to answer questions about planetary motion. It would simply be inappropriate. An example of a small model-inquiry gap (i.e., a good match) might be using a social network model to study individual influence in an organization. Where a broad gap exists, using the model (if it can be used at all) will require vastly more effort, to include extensive calibration if good and comprehensive data exists, than if a suitably simpler model were used. Effective theories act as tools for reducing the gap with lower effort. This is also the reason that recent discussion of DoD’s modeling for defense planning decried over-dependence on detailed models and urged greater emphasis on simpler models (with more detailed models used to study selected issues in more detail, which is often crucial). The lesson is not to choose simple rather than complex, but to have the right family of models for both broad analysis and for in-depth analysis when necessary (Davis, 2014a).
The use of effective theories is a way to give agents some aspects of bounded rationality\(^5\) (Cioffi-Revilla, 2014, p.132) as they try to act in complex environments. With effective theories it is possible to create models and simulations of poorly understood systems that are valuable and valid for a purpose (Zacharias et al., 2008). Comparison of game-theoretic models in Stackelberg games (used as a proxy model for critical infrastructure defense) and empirical data showed that humans deviate significantly but predictably from rational expectations. Failing to incorporate bounded rationality in behavioral models leads to sub-optimal or invalid models (Pita et al., 2010).

**Key Questions:** What are the standard or best-practice approaches for equipping behavioral models with the kinds of heuristic decisionmaking mechanisms that humans demonstrate? When is it important to do so?

Consider the preceding discussion as a discussion on the depth of a model’s theories. What about the breadth of a model? What is the relevant scope of a model? Most systems of interest are sub-systems of larger systems. There will be interactions within the hierarchy of systems. Identifying which interactions and systems are relevant to the modeler’s interest is not always obvious. Casting too wide a net leads to large unwieldy models. An overly parsimonious model may be too incomplete to be valid for purpose.

Herbert Simon argues that most systems are nearly-decomposable federations of subsystems (Simon, 1996; Simon, 2002). This is because there are evolutionary fitness benefits to abstracting away functions into subsystems, rather than having a single, entangled monolithic system of functions (Simon, 2002). Near-decomposability tends to improve adaptability and reduce fragility. This suggests that that composite social-behavioral models need to be well designed so that they have the benefits of near-decomposability.

We are much less sanguine about the role of federated models than was the NRC report (Zacharias et al., 2008). Model composition will be important for particular purposes, but the tendency to adopt “approved” model federations with “approved” input data should be fiercely resisted because the problems of interest tend to be dominated by uncertainty. Analysis in such cases needs to be uncertainty sensitive.

**Key Questions:** When should do federated models be encouraged and, in those cases, how should they be designed and used?

### 3.4 The Scope of Data for Behavioral Models

What about the system data? System data here refers to measurement of the system’s state. The typical roles of system data in modeling are tuning models, validating models, generating

---

\(^5\) Herbert Simon’s bounded rationality refers to making decisions under constraints like imperfect or incomplete information, limited or imperfect computational capacity, and limited time—i.e., the kinds of constraints that almost always apply in real life.
hypotheses of underlying systems behavior, and finally for establishing the appropriate inputs for a given application of simulation. Systems data in our context includes behavioral or social data.

The emerging “Big Data” ecosystem is creating a steady stream of behavioral data (Ohm, 2010). A significant body of data-driven work exists on mobility behavior (De Montjoye et al., 2013), financial behavior (De Montjoye et al., 2015), and personal networks (Sapiezynski et al., 2016). The goal of these is to help make sense of how people behave or make decisions. Such insights would be commercially useful for improving targeted advertising and other applications that deal with transitions from user intent to user action. Much of this work leverages newer data-driven methods to highlight patterns. But they tend to fall short on making sense of the patterns they find.

This sense-making is important if behavioral models are to reach their potential. Social and behavioral simulations will need to bridge the theory-data gap. Characterizing behavior without a theoretical frame is often not enough to determine proper interventions. And positing theories of user behavior without empirical support is not enough.

Key Questions:

Which data-driven methods are most useful for sense-making? Which methods have limited sense-making value (deep learning)? What would a research program focused on developing methods for sense-making look like?

Some of the issues have been discussed recently in a volume summarizing results of DoD’s Human Social, Cultural, and Behavioral Modeling program (HSCB) (Egeth et al., 2014) (see, e.g., Chapter 10).

Causal inference is another consideration that motivates the need for a theory frame around behavioral data. Good causal theories, however, must include variables and relationships that are often omitted. For example, some theories and formalisms do not include feedback cycles. Also, the variables used for hypothesis testing statistical work are often not very appropriate for causal explanation because they are poor proxies for the “real” variables of interest and because they sometimes lock in a simplistic framework (e.g., linear dynamics and trivialized rational-actor decisionmaking).6

Key Questions: The current data ecosystem is strongly skewed towards the collection of observational data. What is the state of scalable methods (quasi-experimental methods) for extracting causal relationships from observational data? Are the limits of causal inference on observational data going to diminish with a larger ecosystem? Or are these limits fundamental? Are there robust scalable ways of eliciting causal insight from experts?

6 Attempting to use quantitative social-science “theories” to inform policy-level issues is strongly criticized in (Davis, 2011, p.326ff), drawing on criticisms from within the social-science community itself (Sambanis, 2004; Kalyvas, 2008). A major problem was said to be the tendency for the quantitative work to be aggregated and too little informed by factors known to be important from more micro-level case studies.
3.4 Bridging the Theory-Data Gap

3.4.1 Initial Observations

Models are sometimes starkly dichotomized models as theory-driven vs. data-driven, but it is better to think in terms of a spectrum.

- On one end, the theory-driven approach is said to rely on *deductive reasoning* – using general system laws to infer behavior in specific instances.
- At the other end, the data-driven approach is said to rely more on *inductive reasoning* – using a collection of specific observations to abstract out governing laws and relationships.
- *Abductive reasoning* is a somewhat intermediate mode of reasoning. Abductive reasoning starts from a portfolio of general hypotheses or laws and proceeds to rank them based on how well they match observed data. Abduction is concerned with making “inferences to the best explanation.” This bridges the gap between theory-focused and data-focused approaches.7

The approach we urge is shifting the balance in social-behavioral search from a statistics-and-data-driven dominance toward a healthier abductive approach in which theory informs data analysis and data analysis informs theory in a continuous dynamic tension. This corresponds as well to urging a relatively greater emphasis on *causal* modeling rather than correlational work.8

Thinking now of modeling and simulation, an agent reasoning in an abductive mode would observe state and have a number of hypotheses to *explain* it. It would then act using a hypothesis (or theory) that is as simple as possible for explanation, but no simpler [an expression of Occam’s Razor attributed to Einstein]. The first part (as simple as possible) encourages models that are more likely to be generalizable to cover new observations (Pearl, 1978). The second part (but no simpler) is important because, especially when data is incomplete and imperfect, it may be very important to persist in using some model complexities that are rooted in knowledge or persuasive theory.9 Data may need to “catch up.” This full process, from observations to best

---

7 Charles S. Peirce discussed abductive inference (Peirce and (ed.), 1940)
8 We recognize, of course, that many papers that have been written about how, allegedly, system models make the concept of causality untenable, how “everything” is correlational because we don’t know the ultimate “true” theory if it exists, and so on. This is not the place to discuss such matters.
9 Some 20th century physicists noted candidly that their theoretical work, the stuff of their Nobel Prizes, sometimes was driven by the “beauty” of the mathematics they articulated, rather than at-the-time empirical information (Dirac, 1939; Weinberg, 1994). At a more mundane level, some aspects of social-behavioral theory are intuitively very persuasive. If so, they should not be omitted without strong empirical disconfirmation. Interestingly, in DoD’s
explanation, is a better description of what scientists do when they seek out governing laws of nature. The less structured process of such abductive reasoning that draws on both data and theory has inhibited its automation.

**Key Questions:** Are there modeling approaches that enable flexible exploration of alternative explanations of observations? Which analytic techniques better represent the abductive process? What could it mean to foster a capacity of programmable abductive model building?

The preceding sections discuss criteria for evaluating social simulations. It may be useful to examine currently implemented social simulations through the lens of these criteria. These simulations are rudimentary, but they can serve as useful caricatures of how to use data to train simulation models in a theory-informed manner.

### 3.4.2 Example: Modeling Belief Transmission: Memes and Related Issues at the Micro Level

The transmission of beliefs is a social behavior of considerable interest. Belief here refers to a unit of imitable behavior, socially shared cognition, or culture more generally. Dawkins called such a unit a meme in his 1976 book, subsequently reissued (Dawkins, 2016). Memes include information, conspiracy theories, aspects of individual identity, and even language. The flexible and structurally valid model of belief transmission can have significant policy relevance for questions about the transmission of health, voting, radicalization, & other types of behaviors.

Behaviors and incentives for behaviors can evolve spontaneously over time and in response to an individual’s information diet. An important explanatory model of belief transmission sees an individual’s or agent’s beliefs and actions as a function of the beliefs and actions of members of the agent’s social networks. Social behavior or culture propagates through social networks. Every person belongs to multiple networks, but infectiousness of such memes (ideas, behaviors, or idioms that spread within a culture) depends on the network in which an agent encounters the meme. Teenagers are more likely to adopt behaviors observed in their peer network. We can illustrate the phenomenon on a simple model based on social networks equipped with a provisional theory of transmission.

Our simple model draws on theories of belief fixation and the intersectional nature of identity. Consider a model that is concerned with only a single binary belief. Belief fixation measures how strongly an agent’s belief state resists changing when confronted with new information. We represent doubt about the belief using Bayesian priors. Fixation refers to how strongly the agent resists changing its beliefs in light of new information. Inquiry, is the process of modeling of combat, many operations researchers have often omitted the factor of morale because it is difficult to measure and “subjective.” Senior officers, historians, as well as some modelers have recognized such practice as absurd.

---

10 Peirce’s conception (Peirce, 1877; Peirce and (ed.), 1940) was that an agent’s main goal was certainty with doubt being repugnant.
by which the agent moves from doubt to certainty. We operationalize this using the concept of intersectionality: an agent resolves doubt by examining the beliefs of neighboring agents in a set of personal networks (e.g. family, friends, workplace, & community). The strength of an agent’s impulse to switch belief is a weighted function of the fraction of each such network with belief contrary to the agent’s. This models Peirce’s “appeal to authority” mode of changing belief; the authority here is the weighted plurality of beliefs in an agent’s local networks. It also reflects the function of social influence in opinion formation. In a simple agent-based simulation, then, an agent’s belief regarding the issue in question (the meme) depends on its prior belief one timestep earlier and the beliefs of the other agents in his network.\footnote{For the illustration, we assumed that each agent was a member of four networks, as shown, the networks being of different character. This particular simulation shows a particularly rare belief/behavior (initial prevalence at 1%) spreading through the population rapidly and then randomly oscillating in frequency around 45% (Figure 3). The illustrative simulation shows the meme “caught on.” This is the result of the parameter values assumed in the model. With other choices, the meme would never catch on, or might propagate for a while and then dissipate again. Thus, even this very simple model can—by varying parameters—generate quite a range of behaviors. Conversely, if behavior is observed, then the model parameters can be inferred.\footnote{The illustrative model is an example of Markov Random Field (MRF) equipped with Bayesian temporal update dynamics. More specifically within MRFs, it is an Ising model (Brémaud, 2013).}} If the initial beliefs of all agents are specified and the simulation is then executed, the fraction of the agents believing the meme will change over time as illustrated in Figure 2.

For the illustration, we assumed that each agent was a member of four networks, as shown, the networks being of different character. This particular simulation shows a particularly rare belief/behavior (initial prevalence at 1%) spreading through the population rapidly and then randomly oscillating in frequency around 45% (Figure 3). The illustrative simulation shows the meme “caught on.” This is the result of the parameter values assumed in the model. With other choices, the meme would never catch on, or might propagate for a while and then dissipate again. Thus, even this very simple model can—by varying parameters—generate quite a range of behaviors. Conversely, if behavior is observed, then the model parameters can be inferred.\footnote{The illustrative model is an example of Markov Random Field (MRF) equipped with Bayesian temporal update dynamics. More specifically within MRFs, it is an Ising model (Brémaud, 2013).}

In equation terms,

$$p_k(t) \sim \text{Beta} \left( \alpha_k(t) \cdot \bar{w} + \lambda \cdot \delta(X_k(t - 1)), \beta_k(t) \cdot \bar{w} + \lambda \cdot \delta(1 - X_k(t - 1)) \right)$$

where $X_k(t) \sim \text{Bernoulli}(p_k(t))$,

$\alpha_{k,j}(t) =$ number of agents connected to $k$ in the $j$th network who believe,

$\beta_{k,j}(t) =$ number of agents connected to $k$ in the $j$th network
who do not believe,

$\omega_j =$ relative weight or leverage of belief in the $j$th network

$\lambda =$ inertia of prior belief].
Figure 3 The Multiple Networks To Which Agents Belong
3.4.2.1 Evaluating the Model

The belief transmission model and similar such models can be useful for social and behavioral modeling and simulation. This is an example of a micro-level model, specifically a micro-level agent-based model (ABM). Since it revolves around a single belief, it might prove simplistic when comparing it to full models of human decisionmaking. In some cases, however, it might be useful for exploring and building intuition about belief transmission about a particular meme.

The question of validity is a primary concern for such models. And validity has different dimensions (Davis et al., 2017). The first dimension is the question of structural validity: does the model actually mimic the key processes in the transmission of memes in the real world? The simplicity of the model suggests that it most likely does not capture all relevant features of the real process. It is an open question whether it captures enough of the relevant features. Another dimension is the question of replicative validity: does the model replicate overall behavior under identical initial conditions? The data ecosystem may be becoming robust enough to tackle questions of replicative and structural validity for this model. Social media platforms and Cellular Data Records can provide useful estimates of both multimodal social network relationships and the memes propagating through these networks. Using these data sources could also enable the model to account for dynamic networks and shock events. It is thus possible, in
theory and with careful calibration, to compare meme transmission in this model to transmission in the real world.

The questions of structural and replicative validity are not synonymous for ABMs. An ABM may mimic system processes closely and still fail to replicate observed macro-level system behavior. And conversely, an ABM may replicate observed macro-level behavior based on internal dynamics that are not identical to the true system’s internal dynamics. Such mismatches may be attributed to incorrect/incomplete specification of model dynamics, hidden/latent/unaccounted system variables, or issues of emergence (chaotic or otherwise) more broadly. The question of validity for ABMs is an area of continuous research effort.

3.4.3 Example: Static, Factor-Tree Modeling of Public Support for Terrorism

Behavioral models may also work at a more macroscopic level. Consider, for example, a model intended to examine the factors that promote or inhibit the public support for insurgency and terrorism (PSOT). The macroscopic nature of the question suggests that the effective theories governing the model should be macro-level to achieve a low model-inquiry gap. A large 2009 DoD study drew on a comprehensive review of social science relating to counterterrorism to construct a composite qualitative model in the form of a “factor tree” (a kind of primitive but broad static causal model) of, e.g., public support for terrorism (Paul, 2009). A later study tested it empirically (qualitatively) and refined it slightly (Davis et al., 2012) (Figure 5). The work was synthetic, across different fragmentary social-science theories.

A next step (Davis and O'Mahony, 2013) went beyond the purely pictorial description to build a computational version of the factor tree, but with numerous degrees of freedom to accommodate different possibilities about how the factors in fact combine. That is, the model incorporated not just different levels of resolution as indicated in the tree, but different structures for the combining relationships. The intent was to enrich the ability to discuss causal phenomena at a point in time. The price paid was deliberate suppression of dynamics. Although simple enough to present in a single page, the factor tree integrates a great deal of knowledge—moving discussion away from what the alleged primary reason for public support is to the many factors that affect it, with the relative significance of factors varying from one context to another as expected. The same factors appeared in subsequent case studies (i.e., the qualitative theory had significant generality), but—as predicted—the relative significance of the factors varied with case (Davis et al., 2012).
The computational version of the model allowed generating broad outcome maps showing the circumstances (contextual variables) under which public support would be expected to be very low, low, medium, high, or very high (Figure 6) (Davis and O'Mahony, 2013).
Figure 6 An Illustrative Outcome Map Showing Public Support Versus Five Contextual Variables. Degree of public support is indicated by a cell’s color (or number). All variables have values between 0 and 10, with the points chosen having values of 1, 5, and 9.

3.4.4 Adding Dynamics with Fuzzy Cognitive Maps

(Osoba and Kosko, 2017) extends the simulation capacity of the Davis-O’Mahony model by adapting the factor tree into a fuzzy cognitive map (FCM). FCMs are a type of expert system for capturing and simulating causal knowledge from experts and data. Both factor trees and FCM models score high on important modeling concerns like producing interpretable and multi-resolution representations of systems and concepts. But FCMs have the added benefits of being able to:
- flexibly represent causal dependences (even feedback or cyclic dependence),
- fuse knowledge sources (e.g. experts and data),
- perform data-driven automatic hypothesis generation, and
- simulating static and dynamic what-if scenarios over short or long time horizons
These added benefits greatly increase the capacity for *abductive model building*. Analytic methods like FCMs can useful for bridging the gap between theory and data for modeling and simulation.

**Figure 7 A Fuzzy Cognitive Map Adding Dynamics to A Factor-Tree Model**

Figure 7 shows the graphical depiction of FCM-PSOT, the FCM adaptation of the Davis-O’Mahony PSOT model. The FCM-POST is now directed graph (digraph) as opposed to the tree structure of the PSOT. Cycles are easy to incorporate into digraphs as the figure shows. The digraph’s adjacency matrix combined with an appropriate signal squashing function at the nodes (e.g. the logistic function) allows the analyst to simulate causal progression by single time steps or all the way to convergence to a fixed-point or limit cycle. Progression in time on FCMs is a simple matrix-vector multiplication composed with a nonlinear squashing operation on the output vector.

The FCM’s digraph structure also enables the fusion of multiple maps e.g. from separate experts on the same topic. FCM fusion amounts to weighted combinations of adjacency matrices (augmented or zero-padded if necessary). Figure 8 shows the graphs and adjacency matrices in an example of such knowledge combination. The figure shows elicited maps from two experts on the blood-clotting process known as “Virchow’s triad” combining into one FCM.
3.4.4.1 Evaluating the PSOT Models

PSOT and FCM-PSOT are examples of macro-level models of behavior (compared to the belief transmission model). The relevant entities are populations of people, not individual agents or people. More specifically, the models operate on the population-level prevalence of beliefs and the interactions among these beliefs. The model’s focus on common beliefs makes the validation process more complex. The interaction between beliefs and the more observable actions or events requires careful calibration in this context (and in more general contexts). The existing data ecosystem currently does not (and most likely will not) inform most of the model variables. We would need careful surveys and ethnographies to measure the pervasiveness of the relevant beliefs or factors. Thus tests of validity (replicative or predictive) would be hard to implement especially for the dynamic FCM-PSOT model.

The value of these models is mainly as a tool to characterize and simulate scenarios based on carefully vetted expert knowledge about causal links. The correctness of the elicited expertise wholly underpins the structural validity of both models. The fusion capability for FCMs means probabilistic limit laws guarantee the structural validity of the final FCM if a large number of independent experts present individual FCMs for fusion.

The FCM-PSOT is an example of an AI method (an expert system) that is directly applicable for social and behavioral modeling.
This discussion has outlined key changes and trends in the data and modeling environment since the NRC report of 2011. The methods highlighted focused on the AI/ML domain. The data conversation examined the emerging ecosystem of data streams with content relevant to behavioral modeling. Social media data stands out as a high-leverage stream. It is just one part of the ecosystem though. The extended mind trend (in which individuals’ cognition and mental states leave digital footprints) suggests that the behavioral data ecosystem may be able record formerly unobservable useful signals.

The AI/ML discussion highlighted a series of innovations that could be of value to behavioral modeling. The innovations seem to hold the most promise for unlocking or combining information from complex or heterogeneous data streams (e.g. language, images, video). We demonstrate the value of the FCM AI approach for developing an expert system model of macro-level behavior. This suggests that older expert-system style AI methods may hold promise for behavioral modeling. Alternatively, reinforcement learning may be of value to adaptive or dynamic agent-based models of behavior. And adversarial training may be useful for capturing or simulating patterns that are difficult to articulate.

The rest of the Working Paper focused on more fundamental questions about the interplay between theory and data in efforts to model and simulate human behavior. We identified a series of questions that will need answers if capacity for behavioral modeling and simulation is to grow. We identified questions on:

The Interplay Between Data and Theory:
1. How do we design models that are valid for purpose and representative of the relevant reality when the data and theories are unsettled? What is the right balance of theory and data focus in models? And what are the best practices on weaving the two together effectively?
2. Is the social science research community able to identify which theories of social behavior are effective theories vs. accurate theories vs. just interesting-but-invalid theories? Is there a stable mechanism or framework for making this distinction? How useful is data for this purpose?
3. Are there modeling approaches that enable flexible exploration of alternative explanations of observations? Which analytic techniques better represent the abductive process? What could it mean to foster a capacity for programmable abductive model building?

Modeling Practice:
4. What are the standard or best-practice approaches for equipping behavioral models with the kinds of heuristic decisionmaking mechanisms that humans demonstrate? When is it important to do so?

5. When should do federated models be encouraged and, in those cases, how should they be designed and used? We are skeptical about the broad the value of model federation.

6. Which data-driven methods are most useful for sense-making? Which methods have limited sense-making value (deep learning)? What would a research program focused on developing methods for sense-making look like?

**Causal Inference:**

7. The data ecosystem skews strongly towards observational data collection. What is the state of scalable quasi-experimental methods for extracting causal relationships from observational data? Are the limits of causal inference on observational data going to diminish with a larger ecosystem? Are there robust scalable ways of eliciting causal insight from experts? We believe the need for interpretable models of behavior to guide intervention makes the development of more capable causal modeling approaches key for useful behavioral modeling.

Our discussion ended with an exploration of rudimentary models and simulations of behavior at different levels of abstraction. These served as simple illustrations of the interplay between theory and data driven perspectives on social and behavioral modeling. They also serve as a concrete canvas on which to test our growing sophistication in evaluating behavioral models and simulations.

**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.


Davis, Paul K. et al. (2012), *Understanding and Influencing Public Support for Insurgency and Terrorism*, Santa Monica, Calif.: RAND Corp.

Distribution Statement "A" (Approved for public release, Distribution Unlimited)


Peirce, Charles S., and Butler (ed.) (1940), *Philosophical Writings of Peirce*, Dover Publications.


Rudder, Christian (2014), *Dataclysm: Who We Are When We Think No One is Looking*, Crown.


