Toward a characterization and systematic evaluation framework for theories and models of human, social, behavioral, and cultural processes

Auroop Ganguly[1], Joe Whitmeyer[2], Olufemi Omitaomu[1], Mirsad Hadzikadic[2], Paul Gilman[3], Peter Brecke[4], Moutaz Khouja[2], Steven Fernandez[1], Christopher Eichelberger[2], Thom McLean[3], Cathy (Yu) Jiao[1], and Erin Middleton[1], Ted Carmichael[2], Amar Saric[2], Min Sun[2]*

[1]Oak Ridge National Laboratory, Oak Ridge, TN
[2]University of North Carolina, Charlotte, NC
[3]Oak Ridge Center for Advanced Studies, Oak Ridge, TN
[4]Georgia Institute of Technology, Atlanta, GA

Abstract

A suite of tools and methodologies has been developed for the characterization and systematic evaluation of human, social, behavioral, and cultural (HSBC) modeling and simulation (M&S) systems. State-of-the-art evaluation methods, which are based on empirical and retrospective validation, Monte Carlo type explorations of the hypotheses space, or visual analytics, may not work well for complex nonlinear and adaptive systems, especially when the space of implicit or explicit parameters is large or when validation data are missing, incomplete, and noisy. This paper focuses on the characterization and evaluation of HSBC theories and models with respect to the ability to tease out causal insights, detect and forewarn about possible emergence, predict predominant or interesting behavior, extract dominant processes, and model extreme behavior. Emergence is defined in various operational contexts, and possible trade-offs with predictability are investigated. The ability to generate actionable insights, for example relatively small advance actions which can reduce the possibility of undesirable outcomes, is critically examined. The simulation test-bed comprises a system of differential equations and disparate implementations of an agent-based model across multiple computational platforms. The evaluation methods range from sensitivity analysis and data mining to process modeling, while the metrics range from information theoretic to statistical. The results suggest that data mining and sensitivity analysis of simulation outputs and observations may offer a way to detect precursors of certain types of emergence, a means to extract causal factors and dominant processes, and possibly offer ways to prevent undesirable outcomes through imperceptible, preemptive actions. In addition, the hypothesis that metrics can be developed to compare and contrast simulations, explore complexity-based emergence, short-term predictability, tradeoffs between emergence versus predictability, and extreme behavior, appears to be supported. However, further tests and investigations are required not only with a wider class of theories, models, and M&S systems, but also with additional observed or simulated data and more comprehensive or better-designed performance measures.

****Corresponding Author Contact: Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37831-6085; Phone: (865) 241-1305; Email: gangulyar@ornl.gov***

Credits and Acknowledgments: The first author and several of the co-authors are employed by the Oak Ridge National Laboratory (ORNL), managed by UT Battelle, LLC, for the U.S. Department of Energy under Contract DE-AC05-00OR22725. This manuscript was developed based on funding received from the Defense Advanced Research Projects Agency (DARPA) of the US Department of Defense (US DoD). The views expressed in the abstract are those of the authors and do not necessarily represent the opinions of their institutions or of the US DoD.

1. Introduction

A news article in Science magazine (Bhattacharjee, 2007) discussed how the U.S. military is interested in enlisting the help of multidisciplinary scientific experts to better understand "*how local populations behave in a war zone*." The article mentioned the "Human Social Culture Behavior Modeling" program at U.S. Department of Defense and indicated, through a few anecdotal examples, the types of prior research emanating from multidisciplinary fields that may be considered the state of the art. The disparate opinions about the possible value of HSBC models were well captured in the article through the diverging opinions offered by retired military commanders and through the limited success of similar efforts in the past on the one hand versus recent and promising technological developments and data resources (Fig. 1) on the other. The critical challenges in systematic evaluation of large-scale social science simulations stem from the inherent multiscale attributes of HSBC processes, models, and theory, as well as from the inadequacy of data and case studies for calibration and validation purposes. The multiscale processes range from psychological profiles of leaders and aggregate crowd behavior to the behavior of institutions or organizations, and of ethnic, geographic, religious, linguistic, and racial groups. The need to adequately handle such processes across scales has spawned a wide range of multiscale social theories, which in turn may be competitive or complementary, and hierarchical or integrated. Also, "surprising" or unusual behavior at one scale may indeed be triggered by minor changes or abnormal behavior at another scale.

Given that validation data are missing, incomplete, or noisy and process understanding is imperfect, the state of the art in model evaluation is expectedly not well developed. Alessa et al. (2006) discussed an "all hands" call for establishing a social science community around the area of "*complexity modeling using agent-based models and cyberinfrastructure*." Quantitative social science theories or theories amenable to quantification in the HSBC domain have seen significant activity in last couple of decades (Carley 1986; Coleman 1990; Opp and Roehl 1990; Opp and Gern 1993; Oliver 1993; Petty et al. 1997; Myers 2000; Ajzen 2001; Ostrom 2007). Many of these theories have been embedded within large-scale simulation systems. Although a variety of methods has been proposed and utilized in the literature (Lewis-Beck et al. 2004), large-scale social simulations have tended to veer toward agent-based modeling (ABM) paradigms in recent years. The ABM approaches themselves have developed along multiple lines. Thus, complex "realistic" agents have been utilized to develop fine-scale models (Silverman et al. 2006) for short-term predictive insights, while models with relatively simple agents (Hudson et al. 2008) have been developed to study emergent behavior. Epstein (2007) describes the generation of artificial societies through ABM, while the Synthetic Environment for Modeling and Simulation (SEAS) described in Chaturvedi et al. (2005) attempts to develop realistic and precise models at country levels. Computational frameworks (Railsback et al., 2006) have been developed for agent-based simulations, from demonstration environments (NetLogo: Tisue and Wilensky 2004) to extensible systems (Collier 2003). However, as discussed by Bonabeau (2002), "*ABM is a mindset rather than a technology*," and as explained therein, thinking of ABM as an alternative to traditional differential equation modeling is wrong, because, "*a set of differential equations, each describing the dynamics of one of the system's constituent units, is an agent-based model*." Bonabeau posits that a synonym of ABM would be microscopic modeling, with macroscopic modeling being the alternative. One example of a macroscopic model is International Futures (IFs), which relies primarily on multivariate statistical analysis and what-if scenario planning with space-time aggregated variables (Hughes 1999). However, even as HSBC models are beginning to explode as a consequence of increased data or computational resources, and perhaps enhanced understanding of social processes, commentaries in recent articles like Subrahmanian (2007) make their lack of success in the real world apparent.

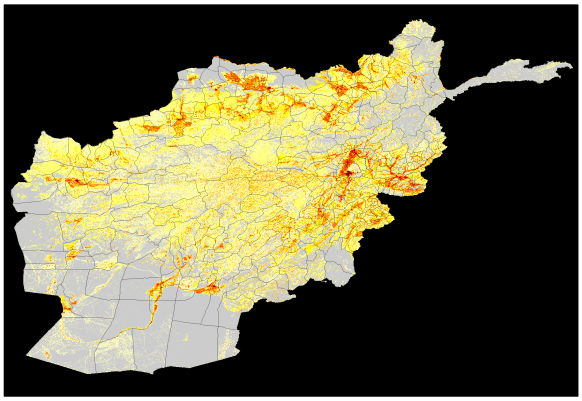


Figure 1: Ambient population count at 1 km resolution in Afghanistan obtained from the LandScan program at the Oak Ridge National Laboratory. The availability of more precise and accurate geospatial-temporal data in recent years is one of the reasons for optimism in our ability to develop HSBC M&S systems that can produce realistic insights for policy makers and military commanders. However, prior failures of such endeavors have led to understandable skepticism and imply a need for caution.

Existing methods for the evaluation of theories, models, and systems relevant for HSBC or similar domains rely on the exploration of the hypotheses (or parameter) space and on empirical validation. These methods include active nonlinear tests of complex simulation models (Miller 1998) as well as structural and parametric sensitivity analysis for the evaluation of complex models (Sterman and Rahmandad 2008). These approaches rely on the design of computational experiments (Santner et al. 2003; Husslage et al. 2006) and empirical validation (Fagiolo et al. 2007; Marks 2008; Windrum et al. 2007). Validation and evaluation in the context of M&S systems for HSBC or similar domains have receibed some attention from multidisciplinary scientific communities (Vicsek 2002; Tesfatsion and Judd 2006; Bryson et al. 2007).

Where models are all too imperfect and validation data are inadequate and noisy, traditional calibration and validation approaches are not likely to succeed. Systematic evaluation of models remains useful however, and is perhaps increasing in importance, as decision makers still need to know how to make best use of the available HSBC process understanding, theories and models, as well as how to utilize available data and computational resources as optimally as possible. In these situations, systematic evaluation may have to take the form of characterization of the space of real-world processes and theoretical simulations. The insights gained may have to be qualitative (e.g., tribal loyalties dominate over individual ideologies in a certain region) or quantitative (e.g., based on structural or parametric sensitivity studies). Retrospective and online analysis of observations and simulations may still be useful where such data are available. However, the ability to characterize the real and simulated worlds may ultimately lead to best-fit recommendations on the selection and use of models. Thus, if the ability to produce emergence is desired for a region where populations are known to adopt ideological positions quickly from neighbors within their social network, the type of models selected may be completely different than those selected for a region where short-term prediction is desired for an ideologically driven culture. In this sense, the characteristics of the real-world processes, model simulations, and desired outputs or insights may all drive the evaluation and recommendation strategy. The strategy may rely on automated approaches like Bayesian or response surface methods, analyst-driven approaches based on matching theory with field insights about the population characteristics, or a semi-automated recommendation, where, for example, guided question-answer sessions ultimately result in the model selection process. The selected theories may operate at multiple scales and/or in a hierarchical fashion. The selection of theories may be probabilistic, and the selection process is likely to directly relate to uncertainty formulations, impacts, and risks.

Key performance indicators for characterization and systematic evaluation of HSBC theories, models, and systems are described next. Following a brief description of an experimental (simulation) test-bed, new results and insights are presented.

2. Key Performance Indicators

Key performance indicators (KPIs) and evaluation methodologies have been developed to characterize and evaluate HSBC M&S systems in the following broad areas:

1. ***Causal Dynamics:*** Extraction of causal insights and dynamical behavior, obtained by combining structural or parametric sensitivity studies with statistical analysis or data mining of model simulations and observations, where available
2. ***Emergent Processes:*** Characterization and prediction of two types of emergent behavior, the first based on behavioral change and the second on system complexity:
    a. <u>Social Behavioral:</u> Interesting and/or significant macroscale social behavioral patterns explained through microscale social theories, which may also be forewarned using statistical change detection and predicted by social models
    b. <u>System Dynamical:</u> System characteristics that make conditions favorable to emergence, measured either as the complexity of microscale dynamics or as macroscale signatures of complex nonlinear dynamics
3. ***Predictability Metrics:*** Measures of predictability, obtained as characteristics of a complex nonlinear system like divergence or nonlinear associations among variables
4. ***Emergence–Predictability Trade-Offs:*** Exploration of the possible trade-offs between emergence and predictability (e.g., in nonlinear dynamics, chaos implies a form of emergence and short-term predictability but longer-term loss of predictability; in agent-based models, complex realistic agents are sometimes thought to produce better short-term predictions even though simpler, interacting agents may model surprises better)
5. ***Dominant Processes:*** The extraction of dominant social processes from potentially massive simulations, as well as possibly limited and noisy observations, through sensitivity analysis and/or data mining approaches like clustering or classification
6. ***Extreme Behavior:*** A spectrum of statistical distance measures for comparing and contrasting simulations and observations in terms of mean behavior and associations, with a particular focus on the behavior of the extremes and their impacts
7. ***Course-of-Action Analysis:*** The ability of an HSBC theory, model, or system, in addition to the ability of the set of KPIs and methods described earlier, to deliver course-of-action analysis and guidance to end users (e.g., military commanders) with some level of confidence; includes precise predictions, evaluating relatively imperceptible actions in advance that may reduce the likelihood of undesirable emergence later, or the converse, and recommending a set of HSBC theories and models geared to a specific context

2.1 Causal Dynamics and Sensitivity Analysis

The need for sensitivity analysis and for the ability to tease out causal insights from simulations has been discussed in the literature (Sterman and Rahmandad 2008). Sensitivity analysis followed by mining of simulation outputs can uncover the spectrum of causal behavior that individual social theories and their combinations are capable of simulating. Mathematical formulations for sensitivity analysis are relatively well understood (Doubilet et al. 1985; Kleijnen 1998; Saltelli 2004; Hazen and Huang 2006). A promising approach is the generalized likelihood uncertainty estimation (GLUE) concept, originally proposed by Beven and Binley (1992) and described briefly in Saltelli (2004, 173). The GLUE approach can be regarded as a simplified implementation of a Bayesian approach and includes a likelihood-like term that assigns weights to various regions of the parameter space. One advantage is the utilization of the Shannon entropy to keep track of the information content of the weights in the parameter space. The weights assigned to model parameters can be used to develop an ensemble of model outputs at any point in time, thus yielding the output PDF. The weights assigned to multiple competing or complementary models may also be considered as parameters, resembling the multimodel superensemble (Krishnamurti et al. 1999) framework developed for weather forecasting.

2.2 Emergent Processes and Emergence Characterizations

Epstein (2007) argues for an operational definition of emergence rather than what he calls "imprecise and possibly self-mystifying terminology of emergence or supervenience." Emergence in complex adaptive systems may appear in unexplained and seemingly inexplicable ways. However, such emergence may have limited value to end users unless it can be interpreted and analyzed for causation. In the context of this paper, we discuss two forms of emergence that are meaningful and usable but not necessarily comprehensive or mutually exclusive. The first is based on social behavior and the second on system complexity. Walker and Smith (2001) discuss the relative deprivation theory, which forms a cornerstone of social-mobilization-based emergence as discussed later in this paper, while Dessalles et al. (2007) and Boschetti et al. (2005) discuss emergence as a property of system dynamics and complexity.

2.2.1 Social Behavioral Emergence

A set of theories for the emergence of social movements is presented (Fig. 2), which includes a treatment of social mobilization based on relative deprivation.

1. Relative Deprivation Theory: Relative deprivation theory asserts that social actors can feel aggrieved or discontented as a result of feeling deprived compared to some reference point. Egoistic relative deprivation is that felt by individuals; fraternalistic relative deprivation is that felt by members of a group about their group. One secondary step or formulation of relative deprivation theory asserts that social movements can arise when fraternalistic relative deprivation passes some threshold.
2. Forms of Relative Deprivation: The feeling of relative deprivation can occur when (1) the social actor (person, group, perhaps even organization) feels deprived compared to some other peer social actor(s); (2) the social actor feels deprived compared to the actor's circumstances in the past; and (3) the social actor feels deprived compared to its expectation of what its current (or future) circumstances should be.
3. Specific Articulation of Relative Deprivation Theory: An individual experiencing fraternalistic relative deprivation and the accompanying sense of discontent has an increased proclivity to be involved in socially disruptive activity compared to when relative deprivation is absent. Past some threshold of relative deprivation, the individual becomes responsive to and will join socially disruptive behavior (a social movement) if there is one to join. Past yet a higher threshold, some individuals will initiate socially disruptive behavior and thereby increase the likelihood that a social movement will form because of the influence of the initiator on the other members of her or his network (group). A socially disruptive behavior is any behavior that noticeably changes the operation of existing social processes and thus may not be favored by all social groups, especially governments.
4. Network Threshold Theory: This theory is tied to the concept of "tipping point." An individual's decision to participate in a group social behavior depends in part on the activities of those around the individual. [Individuals are embedded in social networks (groups) and respond to the circumstances of those networks.] Beyond some threshold value of the network's (group's) situation, usually in terms of the number of individuals doing some behavior, other individuals will join in doing that behavior. That threshold value interacts with an individual's proclivity to become involved in a social movement.
5. Specific Articulation of Network Threshold Theory: An individual will adopt the behavior of others in her or his social network when the number of others in the network doing that behavior passes a threshold. The value of that threshold for individuals is distributed according to some probability distribution.
6. Instigator Theory: Instigator theory asserts that social mobilization is more likely to occur when an agitator or political leader promotes some behavior by other individuals, and through that agitator's or political leader's influence, the individuals instigate some new behavior.

The instigators will advocate a behavior because (1) they are following outside direction, (2) they are socio-political entrepreneurs, and (3) they have a heightened response to relative deprivation.

7. Specific Articulation of Instigator Theory: The presence of one or more instigators in a group advocating some behavior will, in circumstances that are conducive, stimulate other individuals in the group to do that behavior. In those situations where the instigators respond to a situation of relative deprivation, instigators emerge at different values of relative deprivation. (This value may be distributed according to some probability distribution.)
8. Articulation of a Simplified Theory of Social Mobilization: Consider a situation with at least two social groups, and (1) the members of at least one of the groups are feeling relative deprivation, (2) there exists within at least one of the aggrieved groups at least one instigator, an individual who responds in a heightened manner to relative deprivation, and (3) the individuals can communicate with each other. A social movement can emerge through the following mechanism: at some point, if the relative deprivation increases because of unfolding events, an instigator crosses her or his threshold for instigation and begins a socially disruptive behavior. The social network may then grow. Thus, if an instigator is in contact with individuals who are past their responsiveness-to-follow threshold, and if the source of their sense of relative deprivation is similar to that of the instigator, then they will adopt the behavior of the instigator.
9. Stage Set for Social Emergence: If the relative deprivation continues to increase, other individuals pass their thresholds for becoming responsive to the behavior of the instigator(s) in their network and begin the new behavior. [This assumes no external actor stops the instigator(s) (repression).] If the network or group grows in number, other individuals pass their thresholds for adopting the behavior of their network.
10. A Social Movement Emerges: If those two conditions, especially the second, hold true, the network grows. Eventually the tipping point is reached, and the social movement becomes driven primarily by its network dynamics.

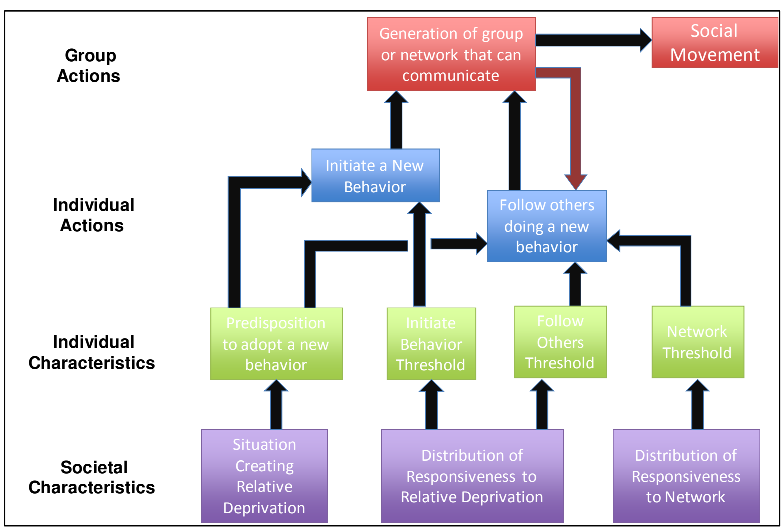


Figure 2: A diagrammatic view of the theory of social emergence. The individual building blocks and their interactions are presented.

2.2.2 System Dynamical Emergence

Emergence has been described and categorized qualitatively (de Haan 2006) and somewhat quantitatively (Deguet et al. 2006). Boschetti et al. (2005) mention that although "self-organization may seem to contradict the second law of thermodynamics that captures the tendency of systems to disorder," the "loss of entropy occurs at the macrolevel, while the system dynamics on the microlevel generate increasing disorder." Boschetti et al. (2005) provide an information-theoretic definition of emergence (Fig. 3) and discuss issues around emergence and predictability from a computational mechanics perspective. Here we investigate the ability to precisely pinpoint microscale dynamics that may be considered sufficiently complex to generate emergence, as well as the ability to extract features from macroscale observables or simulations that are signatures of a system capable of generating emergence. System dynamical emergence includes divergent ensemble runs caused by extreme sensitivity to initial conditions that lead to the presence of strange attractors (see, for example, the Lorenz system: Khan et al. 2005), and also includes cases where relatively simple rules at one scale produce surprisingly ordered behavior at another (e.g., simulating flocks, herds, and schools: Reynolds 1987). Ad hoc definitions of emergence, when a simulation produces seemingly interesting patterns in space and time, may need to be initially developed via spatio-temporal pattern matching and then further investigated to extract dynamical behavior.

Shannon Entropy for Quantifying Possibility of Emergence at Microscale

The Shannon entropy of rules' frequency distribution as proposed by Wuensche (1999) has been adapted for agent-based simulations. The input-entropy at time step *t* becomes:

$$S^t = -\sum_{i=-1,0,1} \left(\frac{Q_i^t}{N}\right) log \left(\frac{Q_i^t}{N}\right)$$

where *N* is the total number of agents and *Q* is the number of agents for each *i* at time *t*. The values of *i* (-1, 0, +1) are those assigned to the behavior variable, *B*, in the social theories. Wuensche suggests that only complex dynamics exhibit high input-entropy variance.

Mutual Information for Quantifying Possibility of Emergence at Macroscale

The mutual information measures the complete dependence, unlike correlations, which are measures of linear associations or rank-based measures that capture only monotonic dependence. The mutual information between agent classes X and Y is the relative entropy between the joint distribution of X and Y [P(X,Y)] and the product distribution P(X) P(Y):

$$I(X;Y) = E_{P(X,Y)}\left[log \frac{P(X,Y)}{P(X)P(Y)}\right] = -\sum_{k=1}^{m}\sum_{k'=1}^{m} P_{k,k'} \, log \frac{P_{k,k'}}{P_{k,}P'_{k}}$$

Figure 3: Following Boschetti et al. (2005), we use two information theoretic measures for complex dynamics that can lead to emergence: (1) the variance of the input-entropy for the microscale at which the rules operate and (2) the lag-1 mutual information, which provides a measure of macroscale dependence structure. These specific formulations should be viewed as illustrative rather than exclusive or exhaustive.

2.3 Predictability Metrics and Measurements

One way to define predictability is to compare retrospectively, or in real-time, with observations and then develop distance measures for the degree of match. However, this approach works only when enough observations are available. If observations are sparse over space and/or time, the HSBC model outputs and states may be validated and updated using approaches like the ensemble Kalman filter (Houtekamer and Mitchell 1998). In situations where observations are inadequate for retrospective analysis or for online updates (e.g., discrete filter formulations) system characterizations may be used to establish bounds on predictability. Thus, linear or nonlinear associations (as determined by, for example, the correlation coefficient or mutual information) between inputs and/or between input-output pairs (where inputs can be a set of input variables and outputs can be a set of simulation variables) would be one type of measure for system predictability. The ability to project the simulations based on simple functional forms, which in turn use current inputs and prior simulation outputs as the functional arguments, may provide an indication of how predictable the system would be with linear or other relatively "simple" nonlinear tools. In cases like the Lorenz series where extreme sensitivity to initial conditions, or "chaos," causes decay in longer-term predictability, the measure of predictability would be based on the sensitivity to initial conditions itself through divergence metrics. Standard methods like Lyapunov exponents and correlation integrals exist to quantify predictability under such conditions [e.g., Khan et al. (2005) estimate a few of these metrics from noisy data and small sample size]. Nonlinear associations, or information content, among inputs and outputs may provide a way to characterize the predictability of a system. Thus, the mutual information, which extracts the information content concerning one set of variables (e.g., inputs) from another set (outputs), can be a measure of the bounds on predictability.

2.4 Emergence versus Predictability Trade-Offs

Within the agent-based model (ABM) community, it has been suggested that ABMs with more realistic and complex agents (as in Silverman et al. 2006 or Chaturvedi et al. 2005), which are inherently more prescriptive, would exhibit more short-term predictability but less ability to model surprising or "emergent" behavior compared to ABMs with simpler individual agents (as in Hudson et al. 2008). These conjectures may be difficult to prove rigorously in the context of large-scale social simulations, ABMs or otherwise, given the differing definitions of emergence, the difficulty of measuring predictability, and the diverse types of models and model complexities. However, we have a clear analogy with complex models in statistics, in which over-parameterization may improve the fit but risks poor generalization. There is also an analogy with nonlinear dynamics, in which additional sensitivity to initial conditions will imply reduced longer-term predictability because the potential space of output grows, even though in the short-term a chaotic system is expected to retain more predictability compared to random noise. Unlike the statistical or nonlinear dynamical analogies, however, balancing the fit and generalization performance, or the chaotic nature of a system with short- and longer-term predictability, may be a significant challenge for HSBC theories, models, and M&S systems. In the latter, the development of equivalents of the information criteria (e.g., the Akaike Information Criteria, or AIC) used to balance model-fit statistics with complexity, or mathematically rigorous formulations to incorporate Occam's razor (MacKay 1995), may not be straightforward. In a similar vein, while the relationship between short- and longer-term predictability, as well as attractor and divergence properties may be well known for chaotic dynamics (Sugihara and May 1990), these insights cannot be immediately generalized to social simulations. Indeed, the trade-offs between emergence and predictability, although seemingly intuitive based on the analogies, cannot be guaranteed as several other factors may overwhelm the process. Nonetheless, there is significant value to understanding and characterizing the trade-off space, if any, as well as in exploring the predictability and emergence aspects independently.

While several methods may be available for characterizing and measuring emergence (depending on the definition) and predictability, information theoretic approaches provide one common platform to explore the possible trade-offs. In fact, Fig. 3 can serve as a surrogate for the emergence versus predictability trade-off, since the microscale complexity measured by the input entropy is a plausible measure of emergence generation, and the mutual information among input-output variable pairs can be a measure of information content and hence predictability. Langton (1991) has explored and interpreted the relationship between input entropy and mutual information in the context of complex dynamical systems.

2.5 Extraction of Dominant Processes

The concept of dominant processes as we define it here has, on the one hand, well-established roots within multiple scientific disciplines. On the other hand, the concept presents certain new challenges in the context of the current problem. The system dynamics literature (Legasto et al. 1980; Ogata 1998; Sterman 2000) attempts to model complex, large-scale problems using a combination of relatively simple approaches that comprise feedback loops, causal links, stocks and flows. The simulation system tries to capture the primary underlying drivers, which are often simple at the microlevel, but through nonlinear interactions may be able to model and simulate complex dynamical processes. The selection and combination of the underlying drivers or processes that can best model a complex system must be inferred from observables and matched with simulations during a calibration process. The computational modeling and simulation (M&S) literature (Zeigler et al. 2000; Kuipers 1994) deals with the mathematical and computational issues of efficient M&S implementation. However, the ability to characterize the underlying processes that form the building blocks of the M&S systems must come either from domain theories and insights or from patterns in observed data or from both.

Once again, calibration and validation of the implemented models require a comparison of simulated and observed data patterns, which may in turn be corrupted with noise. In addition, the relevant observed data may have to be gleaned from massive databases or from relatively limited measured samples. The usable portion of the data gathered from massive databases may reveal only incomplete or partial information, as most of the data may not be relevant for the calibration or validation exercise. The simulated data, while potentially unlimited, need to be carefully generated based on statistical experiments that are grounded in domain and data-dictated insights, and with a view to explore the entire range of plausible outcomes. Thus, the ability to extract broader features or patterns from data, both model-simulated and observed, is a key step. The observed data may be noisy or incomplete, and both the observed and simulated data may have to be extracted from massive but possibly irrelevant data. This is where statistical pattern recognition (Fukunaga 1990) and database mining (Han and Kamber 2006) become important. As our discussions may have also illustrated, while extraction of dominant processes for modeling the real world and the inference of predominant features from data are distinct problems, they share a clear and complementary relation. The other angle for dominant process extraction is through data-guided modeling techniques in which the existence of an underlying functional form can be mathematically proven but the actual description is hidden among what are called latent variables or hidden nodes. Just as examples, artificial neural networks (Haykin 1994) and hidden Markov models (Rabiner 1989) fall in this latter category. Although this broad-brush background may be of value as we extend our ideas in this important line of work, the current problem definition and solution strategy are rather narrow and focused, as explained later in this section.

The inference of dominant HSBC processes from the combined use of observations, model simulated outputs, and domain knowledge (e.g., from a human expert) encompasses several aspects of what could be called knowledge discovery, i.e., abduction, analogy, induction, and deduction (Goertzel 1993):

1. Abduction or hypothesis generation would primarily be under the purview of sociologists, political scientists, and economists, but large-scale simulations can support the process by providing an experimental or simulation test-bed (Magnani et al. 1999). The simulation results, and/or the available theories and models, may guide further data collection or discussions with human experts, all of which can help in testing the hypotheses. Thus, let us consider a fabricated but "realistic" situation where the traditional population (e.g., Pashtuns in Afghanistan) acts according to tribal loyalties in a relatively simple-minded fashion, and for them this "process" dominates over other processes like ideology-driven behavior. On the other hand, let us assume that members of an undesirable group (e.g., Arab terrorists) value ideology over all else and have the power to culturally or economically coerce the local Pashtun population to obey their orders. Consider also the possibility that intelligence reports suggest that the Arab terrorist leadership is likely to attempt to infiltrate the Pashtun. Under these circumstances (which may have prevailed just before and after 9/11 in Afghanistan), if new observations begin to suggest to a U.S. military commander or intelligence analyst that

the local Pashtuns in a region have started to exhibit more ideology-driven rather than loyalty-driven behavior, a hypothesis may be that Arab terrorist leaders have already infiltrated and have started gaining influence. The commander may then run the simulation with suspected terrorist leadership in place, and if the observations at latter times begin to match the new simulations, then the hypothesis may begin to turn to certitude and suggest the need for action.

2. Analogy or reasoning by similarity (Wojna, 2005) would be extremely useful for generalization of models across geographies and time periods, or conversely, to understand the limits of the ability to generalize. This approach would be important when best-fit model or theory recommendations must be made on the basis of an understanding of the dominant HSBC processes. Thus, if a model based on tribal loyalty has performed well in Afghanistan, the same model with tuning of parameters and appropriate changes to the input data may work in another region of the globe, say Mongolia, where tribal loyalties also dominate. However, the model may have limited applicability in regions where ideology- and economy-driven behavior may dominate over tribal loyalties (e.g., Indonesia) or where the type of social interactions may be more of the modern urban type rather than tribal (e.g., Russia around Moscow). Thus, if the population in a region is known to have certain characteristics, or if such knowledge can be inferred from a subset of the observations, then this knowledge can be exploited to design best-fit model recommendation strategies. Automated strategies may include analysis of specific traits or outcomes followed by a probabilistic combination (or triggering/suppression) of theories and models. Manual approaches may include creating a catalog of dominant processes across geographies, time, and other relevant dimensions like ethnicity, religious affiliation, or urban sophistry levels. Semi-automated strategies may include question-answer sessions, in which the computer system asks targeted questions to one or more analysts or domain experts and then prescribes or recommends the best-fit model or model combination. Of course, once a "best-fit" model is proposed and selected by analogy, the choice must be continuously validated under new conditions and accordingly modified.
3. Inductive reasoning (e.g., Tenenbaum et al., 2006) may be necessary when dominant processes must be ascribed to populations on the basis of limited observations or even samples that are known to be biased. Thus, it is likely that in a war zone, information about the population within the control of friendly forces may be collected with relative ease, but corresponding information may be difficult or impossible to obtain in enemy territory. In such cases, the simulations for enemy zones may need to run with characteristics induced from a sample of the population that is within friendly control. The other form of induction that may be useful is Bayesian inference, when simulation parameters may need to be initiated based on prior knowledge, and then estimated and continually updated from new data.
4. Deduction: Deductive reasoning (Evans, 2005) forms the cornerstone of modeling and simulation as well as statistics. Hence this form is implicitly utilized in HSBC M&S systems.

While the extraction of dominant HSBC processes is a challenging but immensely useful endeavor, for illustrative purposes here we define the extraction very narrowly. We test for dominant processes in two ways. First, we utilize parameter and model sensitivity studies to develop an understanding of causal factors and of how they may influence predominant and minority behavior. This is related to the sensitivity analysis and causal insights section as well. Eventually, we envisage data mining of simulation outputs and approaches using a Markov Chain Monte Carlo (Gamerman and Lopes 2006) for exploring the simulation space. Second, we analyze and utilize observed data—even if limited, incomplete, or noisy—and probabilistically assign the data to the processes that may have generated it. The problem definition is simplified in this case by the fact that we consider only processes that have been encapsulated within the modeling and simulation system. The challenge is to classify the relevant observed data, however sparse or noisy, into one or more dominant process/es, which in turn is/are assumed to be the predominant data-generating mechanism/s. A probabilistic classification (Han and Kamber 2006) would apportion the probability among multiple dominant processes. The classification should pit comparable processes against each other. A close analogue in the machine learning or statistical literature is the "mixture of experts" approach (Jordan and Jacobs 2001).

2.6 Extreme Behavior and Distance Measures

Exploration of the model simulation space requires comparing and contrasting simulation results on the basis of a variety of distance metrics, each of which may measure slightly different characteristics of the difference between the simulations. In situations where observations are available, however noisy or incomplete, a similar set of distance measures would be required to compare observed and simulated data. While detailed observations are expected to limit the calibration and validation of HSBC models (which is what makes the characterization and systematic evaluation described earlier very important), the distance metrics are still necessary for situations where such data exist. The space-time nature of the simulation implies the need for geospatial comparison metrics as developed in, for example, Sabesan et al. (2007). The nonlinear relations among variables and processes may require measures of nonlinear correlations (Khan et al., 2007), while the need to understand extremes and tipping points may require measures based on extremes (Kuhn et al. 2007), anomalies, and threat perceptions that consider threshold exceedences (Sabesan et al. 2007).

The first set of metrics includes aggregate measures like normalized mean squared difference (normalized with the product of the standard deviations of the datasets), fractional area coverage (or the total area of the grids with populations that have certain attributes above predefined threshold values) and grid-based difference measures, with or without transformations like the natural logarithm. These metrics may be displayed using traditional error analysis tools. Spatial plots can be developed to visualize the goodness of fit in space and to detect any obvious, relatively large-scale spatial errors.

The second set of metrics comprises spatial auto- and cross-correlation measures as functions of spatial lags. The spatial cross-correlation metrics can also be interpreted as ''spatially aware'' measures of difference. The spatial correlations in each direction can be computed by extending the traditional approaches for calculating autocorrelations and correlations used for time-series analysis (Box et al. 1994; Mills 1991), in the context of spatial data. For two spatially distributed variables that are available in spatial grids [similar to the ''lattice'' data of Cressie (1993)], the spatial dependence structure between the two variables, as a function of spatial ''lags'' (distances measured as multiples of grid spacing) in each direction, can be studied by measuring the spatial linear correlation (Pearson product-moment sample correlation) measure.

The third set of metrics is designed to measure the effectiveness of geospatial data in terms of their exceedence above thresholds and hence the corresponding threat or intensity of disruption they may pose in the context of their end use (Murphy 1993). These metrics combine and refine the concepts of ''equitable threat scores'' and other skill scores used to rank meteorological or climate predictions (Ganguly and Bras 2003). The exceedence-based metrics in Sabesan et al. (2007) are used for evaluating multiple geospatial and geospatial-temporal datasets (e.g., HSBC simulations and relevant observations) rather than prediction skills or signal strengths. The receiver operating characteristic (ROC) curve can be used to visualize the relationship between the false alarm rate and the hit rate. This metric describes the underlying relations in terms of exceeding a threshold and failing to meet the threshold. Measures based on extremes and exceedences, as well as the uncertainties in their estimation, can be related to risk and optimization formulations facilitating end-user decisions.

2.7 Course-of-Action Analysis

The ultimate aim of an HSBC M&S system is to consider the diverse political, military, economic, social, infrastructure, and information (PMESII) aspects of the simulation domain and to provide guidance to decision-makers regarding diplomatic, information, military, and economic (DIME) alternatives. Recent advances in social science theory, data and computational resources, and an interdisciplinary toolbox for complex adaptive systems have spawned a new generation of large-scale social simulations. However, our understanding of social processes remains imperfect, and validation data, where available, continue to be incomplete and geographically sparse. On the other hand, there is a need to balance multiple and perhaps disparate objectives. Thus, short-term predictability must be retained along with the ability to anticipate surprises, while simulation systems must be geographically-aware without sacrificing the capability of generalizing across geographies. Satisfying disparate objectives through a spectrum of competing or complementary modeling and simulation approaches requires a detailed understanding of the strengths and weaknesses of each model and simulation system. In addition, quantitative measures are required to characterize relatively abstract yet critical requirements like the ability to produce emergence or model dominant social processes. A quantum leap is needed to move state-of-the-art approaches for evaluation, characterization, and validation to a stage where they can be used to rank models, suggest optimal model combinations, or recommend best-fit simulation strategies. However, developments in computational and data sciences, as well as insights from other complex domains, suggest new possibilities. Best-fit model recommendations leading to precise predictions of predominant social behavior can help devise strategies to win global wars against terror. The ability of these models to anticipate surprises and their triggers may even help win the peace through imperceptible preemptive actions that reduce the possibility of undesirable outcomes in the future.

The predictive insights generated from HSBC M&S systems can be actionable if and only if end users and decision makers can use the insights to help answer questions like the following:

1. Can our understanding of dominant processes, characterization of performance metrics, and evaluation of model performance lead to improved guidance for decision makers for tactical and strategic use? How can M&S systems be positioned to recommend, whether in manual (analyst-driven), semi-automated, or automated (machine-driven) modes, levers to reduce the likelihood of undesirable behavior?
2. Can the possibility of emergence be detected in advance from minor abnormalities in the data or process? Can emergence of undesirable behavior be terminated in the initial stages by relatively small actions taken early, through root-cause analysis of emergence? Can emergence of desirable behavior be similarly encouraged?
3. Can predominant behavior be predicted with sufficient lead time to enable mitigation actions that can thwart unwanted behavior? Can the prediction of expected outcomes, results of root-cause analysis, or predictive insights on emergence be generalized to time periods, societies, and geographical regions other than those for which they were originally generated? Can the points of major upheavals or reversals in public opinions and their root causes be predicted?
4. Can a combination of insights based on emergence and prediction be utilized to design a multipronged decision strategy? Can the modeling system be used by analysts, or can automated methods be designed, to recommend an action strategy? Can the optimal application of such strategies be pinpointed in geography and time with sufficient lead time to plan and execute necessary actions?

A platform for systematic evaluation must be able to answer the above questions effectively.

3. A Simulation Test-Bed

A prototype experimental test-bed was designed to evaluate the KPI metrics and methods. A system of simple ordinary differential equations generated complex nonlinear dynamics. Three ABMs were used, and the behavior of each agent was determined by maximizing a utility function. The simplest ABM executed on a small-scale environment considered loyalty and ideology alone. A complex ABM built on a desktop environment considered aggregate data and models. The third ABM utilized high performance computing and fine-resolution data/models.

Three leadership theories, specifically legitimacy, representative, and coercion, were implemented by assigning appropriate weights to each factor in the utility function. Neighbor interactions were modeled by using two social mobilization theories: (1) social influence and (2) resistance to repression. These are somewhat different treatments of social mobilization compared to the relative deprivation approach described earlier. Four learning theories, each implemented for change in support for a leader or change in ideology, were developed: socialization, homophily, results-based, and cognitive dissonance. Ninety-six combinations of theories resulted from the nine theories (three for leadership, two for social mobilization, and four for learning or psychological change, where each of the last four can be implemented for leadership or ideology change). The ninety-six theories were implemented in ABMs along with various heuristics for a case study of Afghanistan. The complex ABM solution was also implemented on a larger-scale computational platform with finer resolution data and models. Although the solutions utilized identical theories, the fine-grained implementations required slightly different heuristics. The rest of this section is cursory; the reader is referred to DARPA Foundation Team (2008) for details.

3.1 The Social Theories

Individual behavioral choices were modeled with one of two utility functions (Coleman, 1990):
Cobb-Douglas: $U = (1-L)^{w_L}(1-C)^{w_C}(1-I)^{w_I}(1-E)^{w_E}(1-V)^{w_V}(1-F)^{w_F}(1-R)^{w_R}$
Least Squares: $U = 1 - w_L L^2 - w_C C^2 - w_I I^2 - w_E E^2 - w_V V^2 - w_F F^2 - w_R R^2$

The utility functions encompass seven factors: L is the agent's loyalty to the leader ($L \in [-1,1]$), C is the coercion factor ($C \in [-1,1]$), I is ideology ($I \in [-1,1]$), E is economic welfare ($E \in [-1,1]$), V is security against violence ($V \in [-1,1]$), F is the influence of "close" associates (geographic or social proximity), and R is repression and social influence for defying repression ($R \in [0,1]$) given as *max(0,(-sign(A)*sign(B)))*max((*A^2*-<B*>*^{2}*),0)*, where A is the repressive activity in the area and <B*> is the average behavior of agents within a certain region of the focal agent (a larger region than for influence). The weights are required to be non-negative, to be less than or equal to 1, and to sum to unity ($w_L+w_C+w_I+w_E+w_V+w_F+w_R = 1$). The overall computational goal is to identify the behavior value (B) that would allow a citizen agent to maximize the value of her utility function. The seven components considered here are: For Loyalty: $1\text{-}L=1\text{-}\eta_1\,\text{abs}(O\text{-}B)/2$; Coercion: $1\text{-}C=1\text{-}r*\text{abs}(O\text{-}B)/2$; Ideology: $1\text{-}I=1\text{-}\eta_2\text{abs}(P\text{-}B)/2$; Economic Welfare: $1\text{-}E$; Security from Violence: $1\text{-}V$; Influence: $1\text{-}|B\text{-}\langle B\rangle|/2$; Repression: $1\text{-}R$.
Here B is the considered behavior of the agent to be optimized through the utility function, and O represents an order by the leadership. A variable (e.g., Loyalty: L) is reflected in Cobb-Douglas through the component (e.g., I-L). η_1 and η_2 refer to the agent's support for leadership/ideology, r to the leadership's resources, P to agent's ideology, E to economic dissatisfaction, V to the agent's dissatisfaction with the security situation, and <B> to the average behavior of agents within a certain region of the focal agent. Learning theories are implemented by making each of the variables functions of an agent (i) and time point (t), multiplying the utility function by a learning term ($\lambda_i(t)$), and allowing both λ and P to be "learned" over time in a prescribed manner.

The ideas of Berger et al. (1998) and the "charisma" concept of Weber (1968 [1922]) are encapsulated by the legitimacy theory, which posits that a citizen agent wants to follow the orders of the leadership simply because it is the legitimate leadership. The coercion theory, which posits that a citizen agent wants to follow the orders of the leaders because he or she believes the leadership is likely to punish those who do not obey or to reward those who do, can be traced back to Machiavelli (1985 [1513]) and contemporary advocates (Kiser and Linton 2002; Levi 1988) and fits the behaviorist perspectives in psychology (Sidman 2000) and deterrence theory in criminology (Becker 1968; Stigler 1970). The legitimacy and coercion theories can be implemented by assigning high values to the weights assigned to loyalty and coercion, respectively. The representativeness theory, which posits that a citizen agent follows the leadership only if the leadership advocates what the citizen agent otherwise wants, goes back to Marx (1990 [1867]). The theory finds support in Whitmeyer (2002) or Wickham-Crowley (1993) and can be implemented by setting high values for weights assigned to ideology, economic welfare, and security against violence components. Social mobilization theories have evidence (Roscigno and Danaher 2004.; Opp and Gern 1993; Calhoun 1994) and support (Coleman 1990; Oliver 1993; Ostrom 2007).

The social influence theory proposes that a person is more likely to join collective action the more his or her immediate friends do (McPhail 1991, 1994). The theory can be implemented by setting the weight assigned to influence to a high value and correspondingly reducing the weights applied to the three leadership theories. The repression theory posits that (1) people are less likely to join a collective action the greater the local presence of forces to repress it and (2) in the presence of some repression, people are more likely to join collective action the larger the proportion of people in the observable population who are engaged in this action. The ideas are supported by Opp and Roehl (1990), Opp and Gern (1993), and Marwell and Oliver (1993). The theory requires a high value for the repression weight and lower values for the leadership theory weights.

The population's ideology and support for the leadership and the ability to change these have long been considered important operationally, as reflected in the goal of "winning hearts and minds," and this is reflected in the psychological change or learning theories. In addition to general support for this behavior (Ajzen 2001; Petty et al. 1997), specific support for leadership (Berger et al. 1998) and ideology (Nisbett and Ross 1980) has been shown. Cognitive dissonance theory (Festinger 1957; Nisbett and Ross 1980) posits that an individual resolves a contradiction between a behavior and a belief or attitude by changing the belief. The theory is implemented in the simulation by assuming that the agent's support for a leader increases if the behavior tends to mirror the orders and vice versa. The results-based theory (Coleman 1990) posits that an individual's beliefs and attitudes shift in response to how well he or she is doing. The theory is reflected in the simulation through a shift in ideology in the direction of improved well-being for the agent. The homophily theory (Bourdieu 1984; Carley 1986) posits that an individual's beliefs and attitudes become more similar to the beliefs and attitudes of those close to and similar to the individual. This is an attitudinal version of the effects of social influence seen for behavior in the social mobilization theories, and it has similar empirical support (Opp and Gern 1993). In the simulation, an agent shifts his or her support for the leadership or ideology toward the average support or ideology of similar others in the vicinity. The socialization theory posits that an individual's attitudes do not change in adulthood. Theoretically, this can be because attitudes are set in early adulthood, or because they are heritable (Olson et al. 2001), or both. In the simulation, this means that the agent does not change his or her support for the leadership and does not change his or her ideology. This theory can be combined meaningfully with another change theory to effect the idea that the support for the leadership or the ideology may be changeable, but only up to a point.

3.2 Computational Implementations

The Lorenz equations are a system of simultaneous ordinary differential equations, which when integrated with a certain set of parameter values demonstrates "chaos," or extreme sensitivity to initial conditions, and several other properties associated with chaos (Weigend and Gershenfeld 1994). The Lorenz system demonstrates that simple systems can produce complex behavior. It has been extensively used by the nonlinear dynamics community to demonstrate basic principles. In our examples, the integrated Lorenz series was often intentionally contaminated with random noise or seasonality (Khan et al. 2005). The purpose was to illustrate certain concepts through a familiar example in nonlinear dynamics, which in turn may be useful for drawing analogies to characterizations of HSBC simulations. The Lorenz equations are given by the following:
$dx/dt = \beta(y - x)$; $dy/dt = -xz + rx - y$; $dz/dt = xy - bz$
Chaotic behavior for certain combinations of parameter values (e.g., $\beta = 10$; $r = 28$; $b = 8/3$).

A MATLAB (Higham and Higham 2005) system on a desktop personal computer implemented the loyalty and ideology components of the leadership theories and the Cobb-Douglas and least squares utility functions. The other components of the utility function were assigned zero weights. This simple implementation allowed a set of followers to be instantiated with different ideological states such that the effect of leader orders on instantaneous behavior could be studied. The purpose of this simple implementation, which can approximate the fully developed M&S system only at a high level, was to provide a mechanism to test new theories prior to implementation as well as to determine the value of the various KPIs through simple but easy-to-understand simulations. In the reported tests, followers were assigned to leaders either randomly or in a very specific way, and the impacts of theoretical settings and leader orders were examined.

A simulation was set up based on NetLogo (Tisue and Wilensky 2004), which is typically used as a demonstration platform for ABMs. We implemented the social science theories and models for contemporary Afghanistan in a personal computer environment. The system considered five types of agents: Afghan government soldiers, coalition forces soldiers, Taliban, leaders, and citizen agents. The citizen agents supportive of the Taliban were called Taliban helpers, whereas citizens who were supportive of soldiers/coalition forces became soldier helpers. The rest were neutral. The country was divided into six regions, each with multiple "patches" in NetLogo. The purpose of the six regions was to allow multiple leaders for the Pashtun tribe, and to allow each Pashtun leader to have a geographically defined area of influence on Pashtun agents. Therefore, the regions apply exclusively to the Pashtun tribe. Other tribes had only one leader each, and those leaders had influence on their agents across the entire country. The data for agents and their attributes were developed in creative ways. For example, opium production was used as a measure of economic prosperity. A variety of heuristics was used for agent behaviors like geographical movements. The total number of agents was limited to a maximum of about 10,000 by the NetLogo environment, which required that the behaviors of citizen agents be modeled at aggregate levels. The data utilized were an aggregate version of the data used for the Oak Ridge Mobile Agent Community (ORMAC) below.

The ORMAC platform (Potok et al. 2003, 2000) was utilized to develop an HSBC M&S system identical to the one in the NetLogo platform but with fine-grained data and with agents at much higher resolutions. LandScan population data (Bhaduri et al. 2002) and the relevant geospatial methodologies (Dobson et al. 2000) were used to build a synthetic Afghan population and to geo-locate the 31 million agents. A variety of disparate geospatial sources was utilized to develop and map the agent attributes as well as the theoretical settings. Calibration data were obtained at district levels. The combination of a GIS-based platform with an ABM is by itself a significant step forward (Brown et al. 2005).

3.3 Insights and Limitations

3.3.1 Insights

The HSBC M&S system described here demonstrates two major possibilities. First, it shows that multiple social theories at a variety of levels of aggregation can be instantiated in a single model in such a way that they can be compared and combined. Second, it also demonstrates that a variety of data can be incorporated as available and as needed into the model. The two capabilities are shown by the facts that the demo model runs and that it produces realistic effects. Test cases have verified computational tractability, and a survey of global data sources confirmed the availability of resources to build synthetic populations with the wide range of attributes necessary for the many social theories envisioned for the ultimate system. By using all-source information and disaggregating primary databases (by means of indicator data sets generated with geospatial science methodologies), synthetic populations may be developed at the individual levels.

The applicability of this methodology was demonstrated in an Afghanistan case study. The ORMAC and the NetLogo simulation systems independently identified coercion theory as most representative of the ground truth observed in Afghanistan, illustrating that HSBC systems may be able to extract the best-fit theory from data. This effort compared and contrasted two different theory-driven, agent-based modeling approaches and explored how each could be implemented and evaluated. A very high-resolution model and a more aggregated model were implemented using the same synthetic population and the same base set of social theories to explore actions that could halt Taliban-inspired terrorist activities. The aggregated system was able to explore efforts such as force-on-force events and the population‘s reaction to changes in the security situation. Simulations with agents representing individual social atoms were able to observe the development of small populations of pro-Taliban members. The difference in the results from an aggregate and a finer resolution M&S system was shown. An attempt was also made (DARPA Foundation Team, 2008) to geographically correlate the simulation results with terrorist incidents as recorded in the World-wide Incident Tracking System, which shows that calibration and validation may be a possibility. Utilization of the 31–million-agent Afghan model demonstrated the ability to computationally control (and analytically manipulate) a system with the large number of agents that may be necessary to model populations at the individual level.

3.3.2 Limitations

The insights are confined to relatively simple implementations of leadership theories for citizen agents within the utility function framework, with rudimentary learning capabilities. Social phenomena at the geographical and temporal scales considered here result from multiple interacting processes, and only a small fraction of these have been addressed in this analysis. Even within the context of the theories we used, the characteristic time scales for the systems are not well studied, and the equilibrium end states or various simulation stabilization criteria are not well understood. Any insights regarding data sufficiency must be treated with caution, especially because the suitability of data for calibration tasks and the relation of data to state variables require detailed examination. The value of fine-grained resolutions and computing power should be objectively tested. Experience in other domains indicates there is often an optimal level of aggregation at which best results are obtained (Chapman and Winker 2005; Fadlalla 2005; Carpenter and Georgakakos 2006; Reed et al. 2004; Done et al. 2004). The optimal aggregation level may depend on the inherent nature of the processes, the validity of the theories and models, and the quality and quantity of input and calibration data. Finally, any insights regarding emergent or surprising behavior must be interpreted with caution (Bonabeau 2007; Epstein 2007) given the subjectivity of the insights and the diverse definitions of emergence.

4. Systematic Evaluation Results

The ordinary differential equation system generated surrogate data, which were used to show how information theoretic measures explore the emergence versus predictability trade-offs. Mutual information among variables related directly to predictability from nonlinear tools, while emergence was obtained by studying the attractor dynamics. The MATLAB-based ABM system had the advantages of little overhead and fast execution times. Thus, sensitivity studies had short execution times leading to design of a simple test scenario that showcased emergent effects through (1) changes in predominant behavior, (2) precursors of emergence based on analysis and mining of simulated data, and (3) detection of emergent behavior via data abnormalities. The NetLogo-based ABM was used to

- study the influence of social theory on agent behavior using correlation studies,
- evaluate microscale dynamics based on input entropy to determine ability to produce emergence,
- extract macroscale signatures based on the mutual dependence structure to characterize possible emergent behavior,
- correlate the former dynamics and the signatures to processes and theories,
- develop dominant attributes by relating pseudo-observations to a contained set of theories that could have generated those outputs, and
- produce threat scores that measure the closeness of simulation in terms of the ability to balance hits and false alarms in the context of exceedence thresholds.

The results of the ORMAC-based ABM were compared with the results of the NetLogo-based ABM to obtain insights on how the fine-grained data and more resolved process models impacted the final results. The experimental test-bed and the evaluation metrics or methods were developed to show the feasibility of using these approaches in the HSBC domain. In particular, the definition of emergence requires a degree of subjectivity and contextual information. Thus, the definitions or KPI formulations require additional study before they can be generalized to real-world processes or other HSBC models, theories, and domains.

4.1 Emergence versus Predictability in Nonlinear Dynamics

The Lorenz X (time series of X obtained upon integration) is corrupted with different signal-to-noise ratios (SNR), and an online support vector regression as proposed by Ma et al. (2003) is used for prediction. The skill (1/MSE, where MSE is the mean squared error) and 1/MI (where MI is the mutual information) between the original noisy and the predicted signals are computed for each SNR, and the results are shown below. The plots (Fig. 4) indicate that both MSE and MI increase as SNR increases. Mutual information (MI) appears to capture the system predictability. Measures of predictability for small samples and noisy data are in Khan et al. (2007).

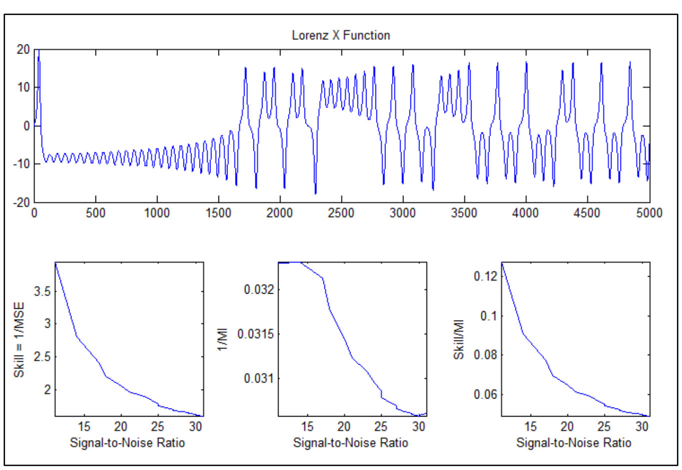


Figure 4: The MI, an information theoretic measure of nonlinear association among variables, relates to the prediction skill (an inverse measure of predictability) in the Lorenz system, which is contaminated with noise. The bottom plots show, from left to right, the skill, the MI, and both.

The Lorenz X, Y, and Z were compared with random and periodic signals in terms of their dynamics in attractor space. The results (not shown here, but see Weigend and Gershenfeld 1994) illustrate the strange attractor. The ability to distinguish emergent dynamics in the attractor space through information theoretic or fractal measures has been demonstrated (Grebogi et al. 1987; Grassberger and Procaccia 1983). However, whether or to what extent the analogies or the intuitions drawn from nonlinear dynamics remain valid for social science simulations is yet to be demonstrated, specifically in light of model imperfections and data limitations (however, see Judd and Smith 2004). On the other hand, the hypothesis that some of these concepts, which can be clearly demonstrated in the simpler models, would remain valid for more complex simulations cannot be rejected outright and may be a good starting point.

4.2 Sensitivity Analysis, Causality and Social Emergence in a Simple HSBC ABM

The MATLAB-based simulation requires selection of four entities. First, either the Cobb-Douglas or Least Squares utility function must be selected. Second, the components of the utility function, or more specifically, the weights attached to the ideology and loyalty components within the utility function (since those were the only two components considered) must be determined. Third and fourth, the impact of ideological positions and the impact of leader orders must be decided. The variable to be studied is the behavior of the citizen agent.

Table 1: Sensitivity analysis and extraction of causal insights from simulation outputs

Number of citizen agents: 100; Number of leaders: 1

–1 = Pro-Taliban; 0 = Neutral; +1 = Pro-Government

P, *L*, *I*, *B* belong to {–1, 0, +1}

Ideological Positioning (*P*)	Selected Utility Function	Leadership Order	Loyalty and Ideology Weights	Behavioral Response
{34%, 33%, 33%}	Cobb-Douglas	+1	75%; 25%	{0%, 34%, 66%}
{34%, 33%, 33%}	Least Squares	+1	75%; 25%	{0%, 34%, 66%}
{34%, 33%, 33%}	Least Squares	+1	25%; 75%	{34%, 33%, 33%}
{34%, 33%, 33%}	Cobb-Douglas	+1	25%; 75%	{0%, 67%, 33%}
{50%, 0%, 50%}	Cobb-Douglas	+1	75%; 25%	{0%, 50%, 50%}
{50%, 0%, 50%}	Cobb-Douglas	+1	25%; 75%	{0%, 67%, 33%}

Table 1 shows several competing or complementary influences on the final behavioral response. Comparison of the first and second rows shows that in certain situations the choice of utility function may not matter. However, comparison of the third and fourth rows shows that the choice of utility function may produce a significant change. Comparison of the first and fourth rows reveals that the choice of weights attached to the different leadership components (loyalty and ideology in this case) may significantly alter the behavioral response even if the same utility function is used for both simulations. Comparison of the first and fifth rows emphasizes the importance of the initial ideological position. In more complex simulations, the ideological position may itself change over time, but this change is expected to be much slower than the change in behavior or mindshare. A comparison of the last two rows once again illustrates the effect of changed weights in the utility function. Detailed analysis of this nature can help explain observed behavior by suggesting possible root causes. Thus, sensitivity analysis and/or data mining of simulations can quantify the change in behavioral response caused by agent initializations, agent decision attributes, and leadership changes. Once new observations or simulations are acquired or generated, extraction of causal insights and precise predictions can be performed. Observed changes in agent decision-making attributes or interactions may lead to predicted changes in mindshare, while observed or simulated mindshare may be explained by changes in the attributes.

As discussed, detailed sensitivity studies based on a thorough statistical exploration of the simulation output space, and perhaps data mining of the simulation outputs, can yield interesting and useful insights that may be hidden in the simulated data. The hidden knowledge, once discovered, characterizes the simulation system and provides guidelines to the end user. Table 2 provides an interesting example.

Table 2: Sensitivity analysis to guide early detection of undesirable behavior

Number of citizen agents: 100; Number of leaders: 1

–1 = Pro-Taliban; 0 = Neutral; +1 = Pro-Government

P, *L*, *I*, *B* belong to {–1, 0, +1}

Ideological Positioning (*P*)	Selected Utility Function	Leadership Order	Loyalty and Ideology Weights	Behavioral Response
{50%, 0%, 50%}	Cobb-Douglas	+1	75%; 25%	{0%, 50%, 50%}
{50%, 0%, 50%}	Cobb-Douglas	0	75%; 25%	{0%, 100%, 0%}
{50%, 0%, 50%}	Cobb-Douglas	–1	75%; 25%	{50%, 50%, 0%}

Table 2 shows the behavioral states of a population that is ideologically polarized into a pro-government faction and a pro-Taliban faction. We have three cases: (1) a leader is pro-government (top row), (2) the leader is neutral (middle row), and (3) the leader is anti-government or pro-Taliban (bottom row). The utility function is Cobb-Douglas, and the weight applied to loyalty and to ideology is always 0.75 and 0.25, respectively. The corresponding follower behavioral states are 50%-50% neutral and pro-government, 100% neutral, and 50%-50% neutral and pro-Taliban, respectively. The seemingly routine sensitivity analysis hides an interesting characterization of this particular simple HSBC system: even if part of a population has an anti-government ideology, a dominating loyalty component will not allow any anti-government behavior unless the followers obey a rebel leader. This can guide an analyst to look for this type of behavioral response to detect the possibility of rebel leadership, or if this response is found, to suggest the possibility of a rebel leader who may be as yet hidden from public view.

We develop a mock scenario to illustrate how the simple MATLAB-based HSBC system, which has been evaluated based on sensitivity analysis and characterized in terms of causal drivers, can be utilized to detect and forecast social emergence and can lead to course-of-action (COA) analysis. The scenario developed here has a strong resemblance to the theoretical construct for social emergence presented earlier in Section 2.2.1, albeit in a proof-of-concept sense:

1. We have two competing leaders, one pro-government and another pro-Taliban.
2. The followers, or citizen agents, are mostly (90%) pro-government ideologically.
3. A few (10%) followers are ideologically pro-Taliban.
4. A pro-government leader has control over 100% of the tribe initially.
5. The anti-government ("rebel") leader secretly emerges and first controls the 10% of the population who are already ideologically inclined towards Taliban.
6. The influence of the rebel leader extends to the pro-government faction, extremely slowly at first.
7. The influence of the anti-government leader spreads quickly among the social network and rapidly builds momentum.
8. Eventually, the tribe is turned to mostly neutral (predominant behavior), with a significant anti-government faction and only a minority pro-government.

The ideological positions of followers and the orders of leaders are pitted against each other within the context of a utility function, and changes in the resulting behavior of the followers are examined as parameter or model selections vary. A mock test scenario is designed in which a strong leader who backs the government gradually yields influence to a rebel leader. The sensitivity analysis results were able to pinpoint barely perceptible indicators of "emergence," (defined operationally as the growth of influence of the rebel leader) from data abnormalities, and to develop predictions regarding the timing for predominant behavior change as well as possible spread of factors conducive to insurgency. This identifies windows where small non-kinetic actions at the right time can preclude the need for stronger military action later. The ability to translate predictive insights to actionable strategies should be further explored with multiple modeling paradigms, commonly available and in-house simulations, and a set of metrics and tools. The results are depicted in Fig. 5.

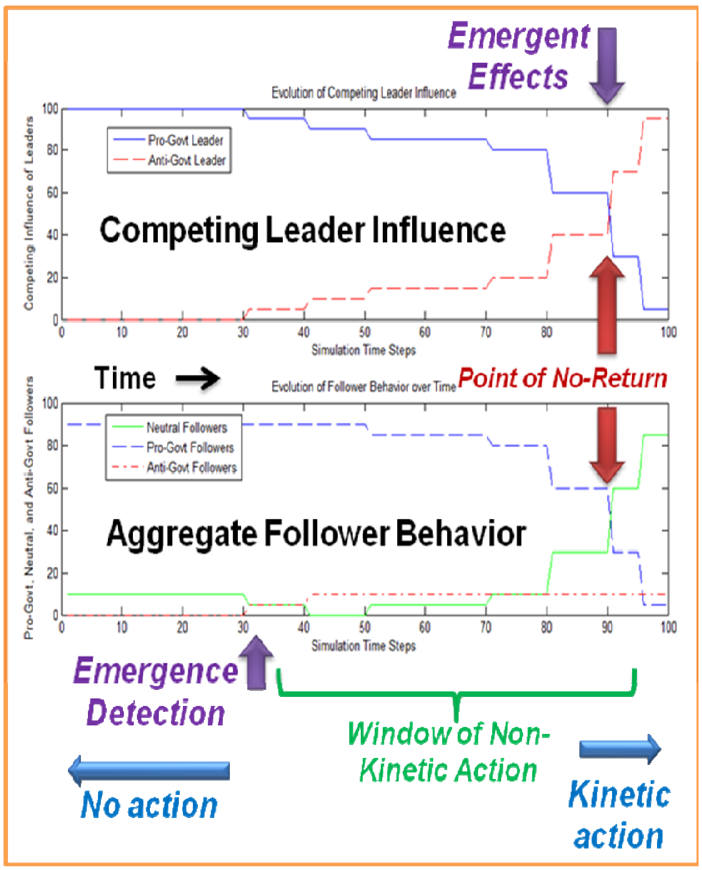


Figure 5: The presence of pro-Taliban behavior, which can happen in this simulation (Table 2) only when a rebel leader is operating, provides advance warning about the possible influence of the leader. In general, this illustrates a case where data abnormalities can provide early warning about emergent behavior in the future. Once the early warning is noted, the simulation can be run to generate the emergent behavior, which in turn can be used to predict the timing and the intensity. Emergence is defined in this case as the point of no return when the predominant behavior changes. This is the point where kinetic action may be the only option. However, prior to this time, a window of non-kinetic action exists when imperceptible action can delay or eliminate the possibility of undesirable emergent behavior.

A MATLAB-based example in Section 4.1 can also be used to illustrate how dominant processes may become evident from sensitivity analysis. Thus, consider the first and fourth rows of Table 1, which show how a switch in the way followers weigh loyalty versus ideology in their individual utility maximization changes predominant behavioral states. In this simple example, assuming all other processes and variables remain constant, the dominant process (the relative influence of loyalty versus ideology) can be easily discerned. In a broader setting, simulated outputs or observations (however incomplete or noisy) may be processed through data mining or information theoretic approaches to characterize dominant social processes that may have generated the observed or simulated data. An example is where simulations generated by a suite of dominant processes are mined and classified into groups, and new observations or simulations are classified (perhaps probabilistically) into one of the groups. This is discussed later.

4.3 KPIs from a Complex HSBC ABM on a Demonstration Platform

The NetLogo-based system was used to develop a demonstration scenario, which in turn was adapted from a realistic event in Afghanistan. This realistic scenario was utilized to characterize and evaluate the system and analyze the outputs in a systematic fashion.

4.3.1 A Storyline and Demo Scenario

Numerous reports in the press in late 2007 and early 2008 concerned the resurgence of the Taliban in many regions of Afghanistan. One typical article addressed the gradual growth of influence of the Taliban in an area close to Kabul, brought about in part by the intimidation or replacement of the leadership (Parenti 2007). We sought to replicate these events with the simulation model and in so doing, to test the effect of the social influence theory.

A demo scenario was generated on the basis of reported changes in a region near Kabul. The leadership in one region ("Region 2") changes its orders from neutral to pro-Taliban (an effect of increased Taliban activity and control in the area). With social influence turned off (weight 0) the proportion of citizen agents in Region 2 who are pro-Taliban becomes instantly large (in one time step) and completely constant. With social influence turned on (weight 0.4), the proportion of pro-Taliban increases fairly steadily to a rough stability in 15 to 20 times steps. All of this occurs for the legitimacy theory setting of weights. When social influence is turned on, they are scaled back 40%.

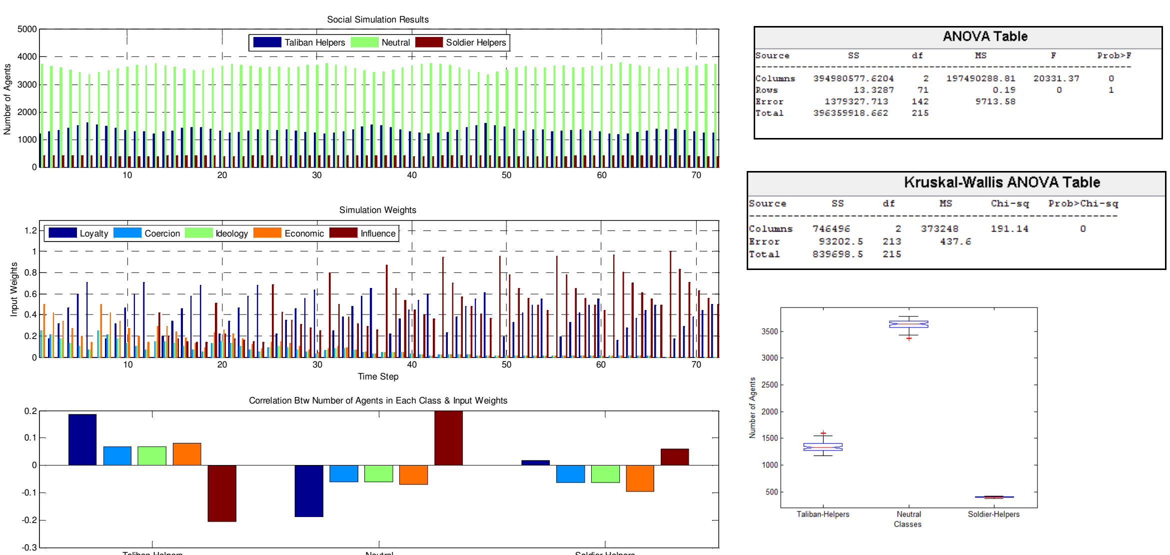
With this in mind, we ran a scenario in which the leadership in Region 2 changed its orders from neutral to pro-Taliban in time step 5. We varied the weights of (1) social influence, to see how its setting affected the final equilibrium, the time to change from one equilibrium to another, and the difference between the low and high equilibria; (2) the radius of the effects of influence, either two patches or three; and (3) loyalty, the strongest weight in legitimacy theory. The idea behind these changes was that the two theories combined in the scenario emphasized those two utility function weights (loyalty and influence), so the data would allow a sensitivity analysis of the effects of those parameters separately and together, as well as the effect of one key parameter of social influence, its radius. We ran the simulation twice, once purely with the legitimacy theory of leadership and once with the legitimacy theory tempered by the social influence theory. Specifically, with only the legitimacy theory, the utility function weights for social influence and repression were set to 0 (utility function weights: loyalty 0.6, coercion 0.1, ideology 0.1, economic welfare 0.2, violence 0, influence 0, repression 0). When social influence was added, the utility function weight for social influence was raised to 0.4, and those for legitimacy theory were reduced by 40 percent (utility function weights: loyalty 0.36, coercion 0.06, ideology 0.06, economic welfare 0.12, violence 0, influence 0.4, repression 0). The psychological change (or learning) theories were turned off. We replicated the empirical events through one simple change. In the simulation model, the initial order from the leadership in Pashtun Region 2 (the region containing Kabul) is for the population to be neutral. Correspondingly, most of the population in this region is neutral. We mimicked the effects of intimidation or replacement of the leadership by the Taliban by changing the leadership order to pro-Taliban in time step 5. The results were as follows. With the pure legitimacy theory and no social influence, the change in the leadership order had an instantaneous effect: in the next time step, the number of Taliban helpers was several times larger and, in fact, dominant throughout Region 2, and the number of soldier (pro-government) helpers was correspondingly much smaller. After that 1-time-step change, the number of Taliban helpers remained constant. With the social influence theory added, the change in the leadership order had a much more gradual effect. Specifically, the increase in the number of Taliban helpers after one time step was small and concentrated in a small area of Region 2. Over the next approximately 15 time steps, the increase in Taliban helpers spread throughout

Region 2 at a fairly constant rate. After 15 to 20 time steps, the number of Taliban helpers remained roughly constant with some fluctuations, at a level somewhat higher than under the pure legitimacy theory. The increase in Taliban helpers did not spread beyond Region 2. Three insights were gained from this exercise: (a) It is possible to replicate at least the broad features of empirical events that fall within the scope of the simulation model; (b) A combination of social theories may yield more realistic results than any single theory. Instantiating the legitimacy theory permitted the change in the orders of the leadership to have effect, but the social influence theory was necessary to add the realism of a gradual spread of that effect; (c) A combination of social theories may produce effects that are not just the sum of the effects from each theory.

4.3.2 The Key Performance Indicators

The NetLogo simulation was implemented for 72 time steps using different input weights; each time step was replicated for 41 runs. The following performance indicators were evaluated:

1. **Causal or theoretical insights:** Correlation of outputs with weights assigned to social theories to determine the influence of the individual theories
2. **Emergence based on microscale dynamics:** Inferring system complexity and the likelihood of emergence from microscale dynamics quantified through the input entropy, with an understanding of the causal behavior based on the social theories implemented
3. **Emergence based on macroscale dynamics:** Inferring system complexity and the likelihood of emergence from macroscale signatures quantified through the mutual information dependence structure, with a quantification of the contributing social theories
4. **Emergence versus predictability:** A common platform for exploring the emergence versus predictability trade-off, with normalized measures for both
5. **Dominant process characterization:** Extraction of dominant processes or attributes from data by matching sample data to the simulations generated from underlying theories
6. **Extreme behavior with a threat measure:** A measure that quantifies the contrasts between simulation or observations in terms of extremes through a threat score



ANOVA Table

Source	SS	df	MS	F	Prob>F
Columns	394980577.6204	2	197490288.81	20331.37	0
Rows	13.3287	71	0.19	0	1
Error	1379327.713	142	9713.58		
Total	396359918.662	215			

Kruskal-Wallis ANOVA Table

Source	SS	df	MS	Chi-sq	Prob>Chi-sq
Columns	746496	2	373248	191.14	0
Error	93202.5	213	437.6		
Total	839698.5	215			

Figure 6: The average number of agents per class (top left) is compared with the weights assigned to the utility function (middle left) and is correlated with input weights (bottom left). The two-way analysis of variance (ANOVA; top right) and two-way Kruskal-Wallis tests reject the null hypotheses that the mean or medians, respectively, of the classes (columns) are equal across time. The p-values and the box-plots (95% confidence bounds do not overlap) support the rejection.

The influence of the theoretical settings (in this case, the weights assigned to various components of the utility function) is clearly shown in Fig. 6 where the difference of the agent classes is seen to be statistically significant as the theoretical parameters change over time.

The typical interpretation of Shannon (or input) entropy is that it specifies the number of bits required to encode (and transmit) the classification of a data item. The entropy is smaller when the data item is more "pure" and is larger when the data item is more "impure." Therefore, entropy has been described as a measure of the rate at which environment appears to produce information. "*The higher the entropy rate, the more information produced, and the more unpredictable the environment appears to be*" (Crutchfield 1994). If entropy is used as a measure of the predictability of classes, then the smaller the class entropy, the more predictable the class would be. For example, if all agents belong to one class, then the entropy is zero, and no bits need to be transmitted because the receiver knows that there is only one outcome (Lee and Xiang 2001); therefore, no uncertainty exists and the class predictability appears much higher. Three classes with the same number of agents will have higher entropy and higher variability because it is more difficult to classify each agent. The input-entropy is a measure of the inherent complexity of the system and hence a necessary condition for complexity-based emergence. Fig. 7 shows the average input entropy for each class and the sum of the average for all classes at each time step, as well as the variance of the input entropy in each class and for all classes.

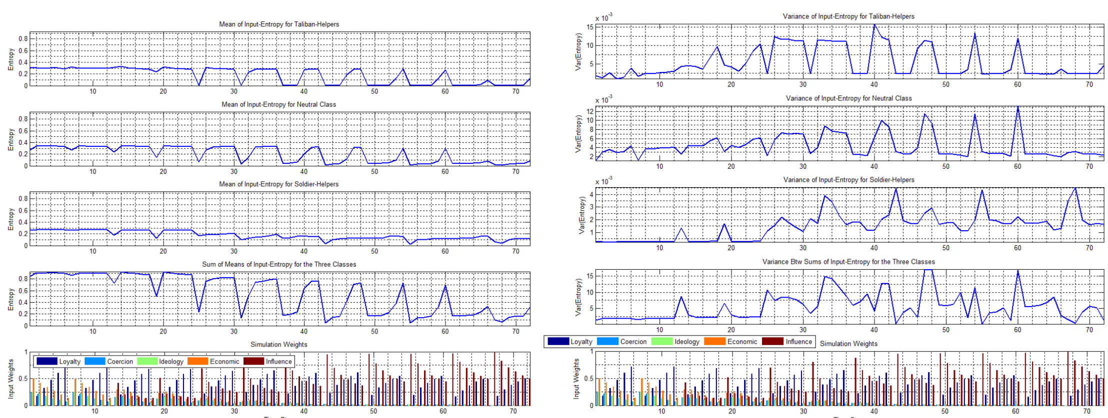


Figure 7: The input-entropy (left set) and variance of input-entropy (right set) plotted for (top to bottom) anti-, neutral, pro-government, and all agents as a function of the input weights (along the bottom). The low variance at the beginning and end of the simulation period points to relative stability, but the period in between is more volatile.

As indicated in Fig. 7, as the input weight for influence increases, the input entropy for each class and all classes decreases. The variance is smallest during the first 20 time steps, is much higher between time steps 20 and 60, and decreases to almost zero at time steps greater than 60. This distribution may further suggest the presence of phase transition (Langton 1991); however, the direction of the transition is not clear from this experiment. The transitions from an ordered state into a disordered state and vice verse (i.e., phase transitions) are sometimes described as indications of emergence (Langton 1991). Of particular interest is that input-entropy decreases when the input weights for influence and loyalty are 0.0 and at least 0.5, respectively. Minimum input entropy is observed when the weights of influence and loyalty are greater than 0.5 and less than 0.5, respectively. Increased input-entropy gives lower variance. The variance is at a maximum when the input weights for loyalty and influence are approximately 0.4 and 0.6, respectively; and it is at a minimum when the input weights for loyalty and influence are approximately 0.9 and 0.25, respectively. The fact that higher interactions (indicated by relatively larger influence weights) suggest a greater possibility for system dynamical emergence appears intuitive. The mutual information is plotted in Fig. 8 with respect to two randomly chosen time steps as reference (in this case, the time-steps 1 and 48). The variance of the mutual inputs provides similar insights as in Fig. 7 earlier, but in terms of macroscale signatures of emergence.

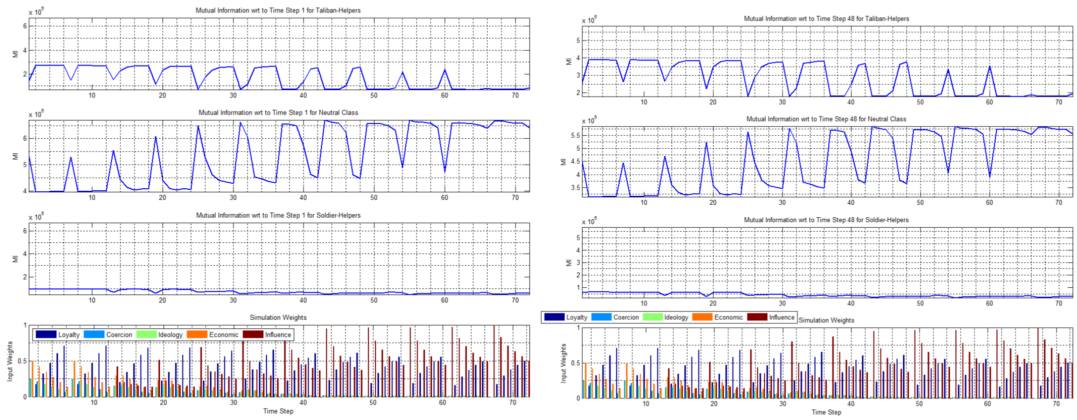


Figure 8: The mutual information, plotted against two random references, exhibits several features similar to the input entropy in Fig. 7. The MI starts at some high/low state with much smaller variability, proceeds through a transition with higher variability, then respectively arrives at a new low/high state with much smaller variability. Statistical tests of significance (not shown) confirm that the means of the mutual information are significantly different for each reference, even though the overall trends look similar and are of more interest.

Fig. 9 shows the lag-1 mutual information, which is a measure of nonlinear dependence among the autoregressive components and hence a measure of predictability in the system. Also shown is the correlation with relative weights assigned to each component in the utility function.

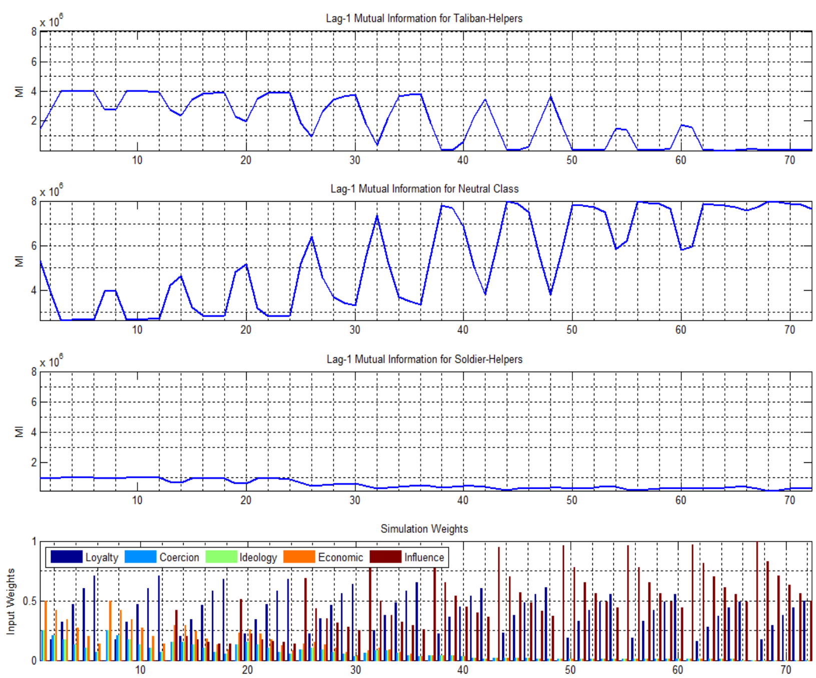


Figure 9: The lag-1 mutual information (MI) for the number of agents in each class. Evidence of a phase transition is observed for agents of all classes including pro-government (soldier-helpers: SH), but seems more pronounced in anti-government (Taliban-helpers: TH) and neutral (N) agents. A trend with simulation time steps is observed: from high to low MI for TH and SH agents and from low to high MI for N agents. This is similar to Fig. 8.

Table 3 shows the correlation coefficients (CC) between lag-1 MI and the relative weights of each attribute for each agent class. The t-test statistic shows that all the attributes are significant at the 0.05 level except for the CC between loyalty and soldier-helpers.

Table 3: Correlation coefficients between lag-1 MI and relative weights (see Fig. 9)

Attributes	**Taliban-Helpers (TH)**		**Neutral (N)**		**Soldier-Helpers (SH)**	
	CC	p-value	CC	p-value	CC	p-value
Loyalty	0.3829	0.0009	-0.3498	0.0026	0.2194	0.0641
Coercion	0.5198	0.0000	-0.5870	0.0000	0.7790	0.0000
Ideology	0.5198	0.0000	-0.5870	0.0000	0.7790	0.0000
Economic Welfare	0.5111	0.0000	-0.5806	0.0000	0.7965	0.0000
Inluence	-0.7337	0.0000	0.7694	0.0000	-0.8527	0.0000

The MI of one class of agents gained by the knowledge of another class of agents is particularly interesting because it measures the mutual dependency between the two classes (i.e., it measures how much the uncertainty about one class is reduced by knowing the number of agents in another class; Cover and Thomas 2006, 19–25). We computed the MI of the number of Taliban-helper (TH) agents due to the knowledge of the number of neutral (N) agents and soldier-helper (SH) agents, respectively, as well as the MI of N agents due to the knowledge of SH agents. It should be noted that MI is symmetric. The plots of these measures are the 3rd, 4th, and 5th plots in Fig. 10.

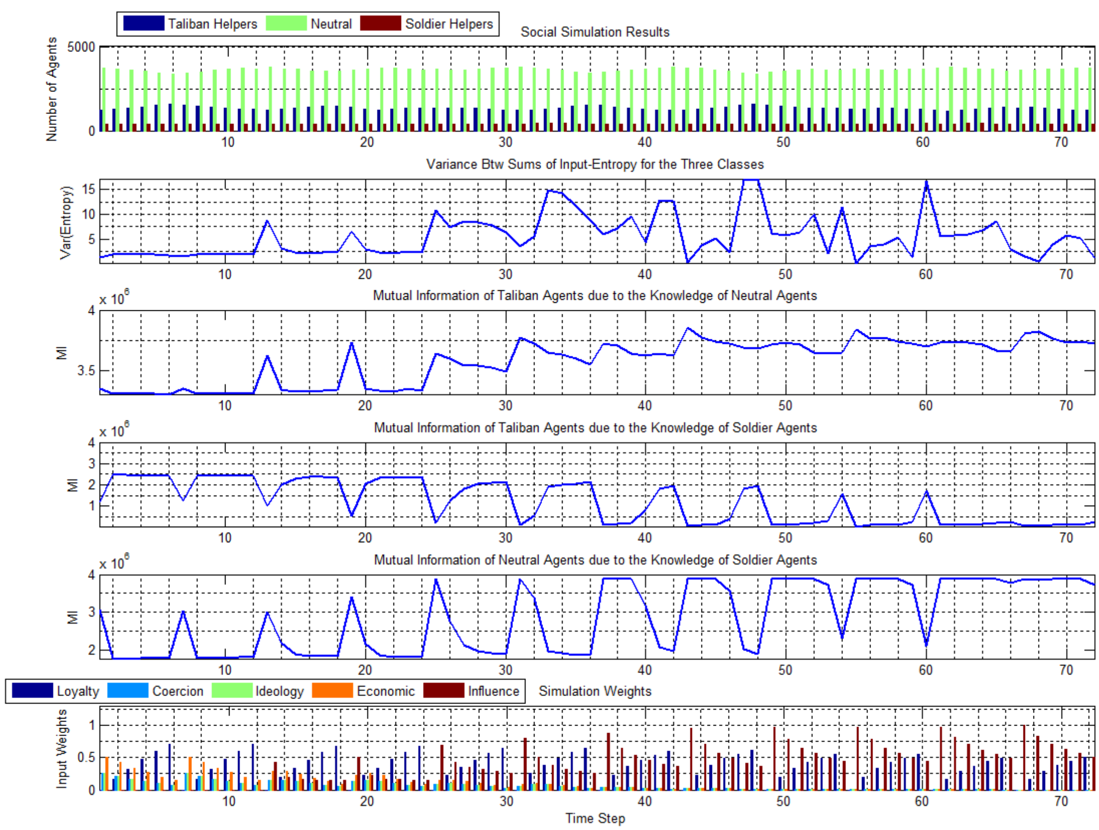


Figure 10: The average number of agents in each class, the variance between the sum of entropy for the classes, the MI of TH agents due to N agents, the MI of TH agents due to SH agents, the MI of N agents due to SH agents, and input weights. The MI values are plotted on the same scale.

The interesting features in Fig. 10 are as follows:

1. For these three plots, MI never equals zero. This indicates that the number of agents in each class pair (for example, TH and SH agents) is dependent.
2. At time step 7, after the leadership was switched from pro-Taliban to neutral, we see a sharp increase in MI for N and TH agents (5^{th} and 3^{rd} plots in Figure 10 respectively) and a sharp decrease in MI for TH agents (4^{th} plots in Figure 10); this is not obvious from the variance of input-entropy plot for the three classes (2^{nd} plot in Fig. 10).
3. A sharp decrease in MI for TH agents is seen whenever either the input weight for loyalty is 0.0 (regardless of the input weights for other attributes) or the input weights for coercion, ideology, and economic welfare are equal and the input weight for loyalty is 0.5 or less. This is an intriguing observation that may help develop simulation experiments.

The behavior of normalized entropy and MI in a scatter plot has been investigated by Langton (1991). If we interpret entropy as a measure of emergence and MI as a measure of predictability, the scatter plot can be interpreted as a trade-off between predictability and emergence. The scatter plot of normalized entropy versus MI for each class is shown in Fig. 11. The scatter plots for both TH and SH agents are convex, but the convexity is more pronounced in the plot for SH than in the plot for TH agents. The plot for N agents is concave, however. Consequently, for TH agents, the maximum MI occurs at about the maximum entropy; for N agents, the maximum MI occurs at minimum entropy; and for SH agents, the maximum MI occurs at maximum entropy.

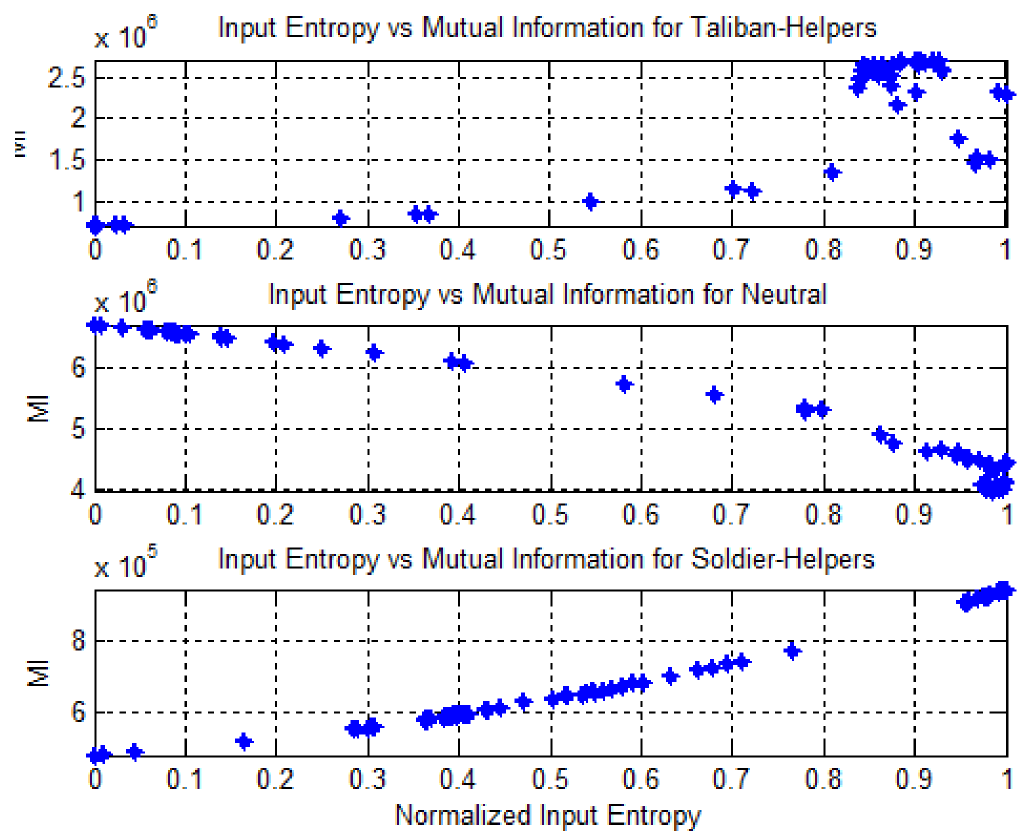


Figure 11: Scatter plot of MI versus normalized input entropy for each agent class, which may be one way to represent the system-dynamics-driven trade-offs between emergence and predictability

Fig. 11 again shows evidence of phase transition by way of the gap between the scatter points (see Langton 1991). These gaps appear to classify the points into two or more clusters, which is a trait that has been attributed to the difference in the orders.

- For SH agents, we have three phases: normalized entropy < 0.2 (one state), normalized entropy between 0.3 and 0.8 (transition phase), and normalized entropy > 0.8 (another state).
- For N agents, we have normalized entropy ≤ 0.4, between 0.5 and 0.8, and > 0.8.
- For TH agents, we observe normalized entropy < 0.1, between 0.1 and 0.5, and > 0.5. This class has more gaps than the other two classes. This may be an indication of some useful information that is not clear from this experiment.

The extraction of dominant attributes from this particular scenario is difficult as the variation among the number of agents within a class is limited and the effect of any of the five attributes is difficult to distinguish because of the nature of the experiments. Future studies with well-designed experiments can improve upon this aspect. However, Fig. 12 shows how the input weights and the number of the agents can be visually analyzed.

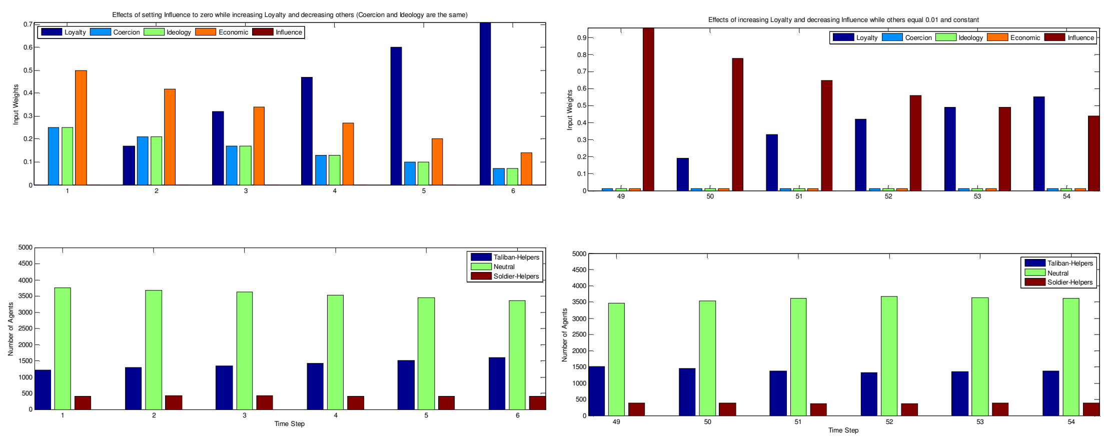


Figure 12: The input weights (loyalty, coercion, ideology, economic welfare, influence) at top and the number of agents per class (TH, N, SH) at bottom for times 1–6 (left) and 49–54 (right)

We note a few interesting points for which we do not yet have full explanations or interpretations. Fig. 7 indicates that all agent classes have minima for entropy (almost zero) at time steps 43, 55, and 68, all of which had loyalty weights of 0.17 or less and had input weights for coercion, economic welfare, and ideology of 0.02 or less. For all agent classes, the variance approaches a minimum (close to zero in some cases) during the first 20 time steps, then takes a much higher value between time steps 20 and 60, and decreases almost to zero at time steps greater than 60. This distribution may further suggest the presence of phase transition; however, the direction of the transition is not clear from this experiment. Notably, the variance is at a maximum for TH, N, and S agents at time step 40, 60, and 43, respectively.

Table 4: Input weights at critical time periods

Time Step	Loyalty	Coercion	Ideology	Economic Welfare	Influence
40	0.45	0.03	0.03	0.03	0.45
43	0.00	0.02	0.02	0.02	0.95
60	0.55	0.01	0.01	0.01	0.44

The significance of the input weights tabulated in Table 4 may include the fact that for time step 40, the weights for loyalty and influence are the same; for time step 43, the weight for loyalty is 0.00; and for time step 60, the weights for loyalty and influence are almost equal and the weights for other attributes are equal to each other.

Distance measures are defined here as metrics that quantify the distance between one simulation output and another or between a simulation output and a set of observations, however sparse or noisy. Thus, correlations, mean squared errors, and information entropy are all useful distance measures. Although simple and commonly used distance measures are useful, here we illustrate a measure that is based on how well the extremes of one align with the extremes of the other. Thus, a mutual exceedence is defined as a "hit" and the reverse is a "miss" and so on. A threat score can then be defined as [hits/(hits + misses + false alarms)]. A class of metrics of this nature has been defined by Sabesan et al. (2007). Usually the measures are employed to compare simulations and observations, but as an illustration, Fig. 13 shows the threat score when one agent class (SH) is used as a base class.

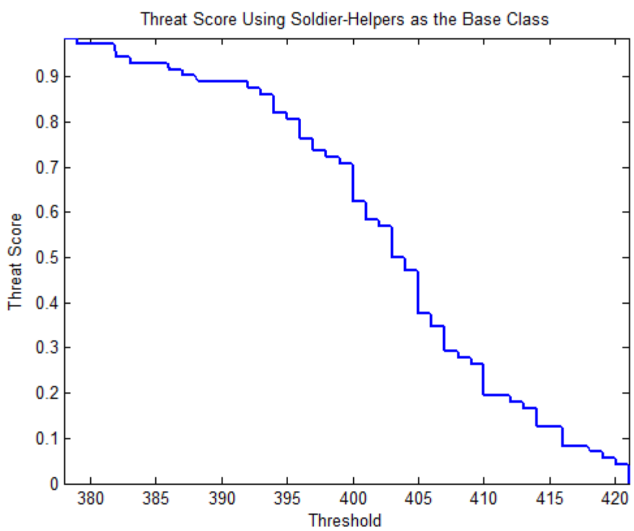


Figure 13: The threat score computed as a function of exceedence of various thresholds, and plotted as a function of the threshold. Measures like these can be useful to understand how datasets, in this case multiple simulations or observations, differ from each other in the context of extremes. This illustration uses an agent class as a base class and compares it with another agent class.

4.4 Comparison of Aggregate and Fine-Resolution HSBC Systems

The ORMAC-based system, with 31 million agents and fine-resolution input data, was compared with the NetLogo-based system, which used a maximum of 10,000 agents and aggregate-level inputs. The initializations were slightly different, with ORMAC using polling data from Gallop for a deterministic initialization and NetLogo using samples from a probability distribution. The geographical and temporal domain was present-day Afghanistan, and the end result was the number of agents with one of three behavioral modes (pro-Taliban, neutral, or pro-government) corresponding to the population mindshare. The social theories embedded in each system were identical, and the test simulations focused on a test of the leadership theories described earlier. Each of the two systems was run nine times (Table 5), with each run instantiating one of three possible leadership theories (legitimacy, L; representative, R; or coercion, C) and one of three time resolutions (3 days, 7 days, or 14 days per "tick," where a tick corresponds to a clock time).

Table 5: The nine simulation runs

Run Number	1	2	3	4	5	6	7	8	9
Days / Tick	3	3	3	7	7	7	14	14	14
Leadership Theory	L	R	C	L	R	C	L	R	C

Because the focus was on leadership theories, five leader assassinations were introduced to explore the effects of disruptive events (Table 6). The events were based on realistic observations in the region. The different simulations provided us with a test set of outputs to demonstrate the value of statistical distance measures. The nine runs are described in Table 5.

Table 6: Events in the simulations

	Event 1	Event 2	Event 3	Event 4	Event 5	End
3days/tick	11	26	27	47	58	68
7days/tick	5	12	12	21	25	35
14days/tick	3	6	6	13	16	26

The aggregate results from the nine simulation runs are shown in Fig. 14. The significant bias in the ORMAC versus NetLogo outputs, even at the end points of the simulations where the outputs at successive time steps appear relatively stable, is obvious from the plots. The only exceptions occur during the instantiation of the legitimacy (L) theory and in that case, only for the number of pro-Taliban agents. The fact that simulations differing primarily in their spatial resolutions result in such large relative biases is cause for concern. Comparisons such as these may help improve model outputs, in this case to correct bias errors. The other interesting aspect is that the ORMAC outputs appear more responsive to the disruptive events, while the NetLogo outputs exhibit more random fluctuations but less response to the events.

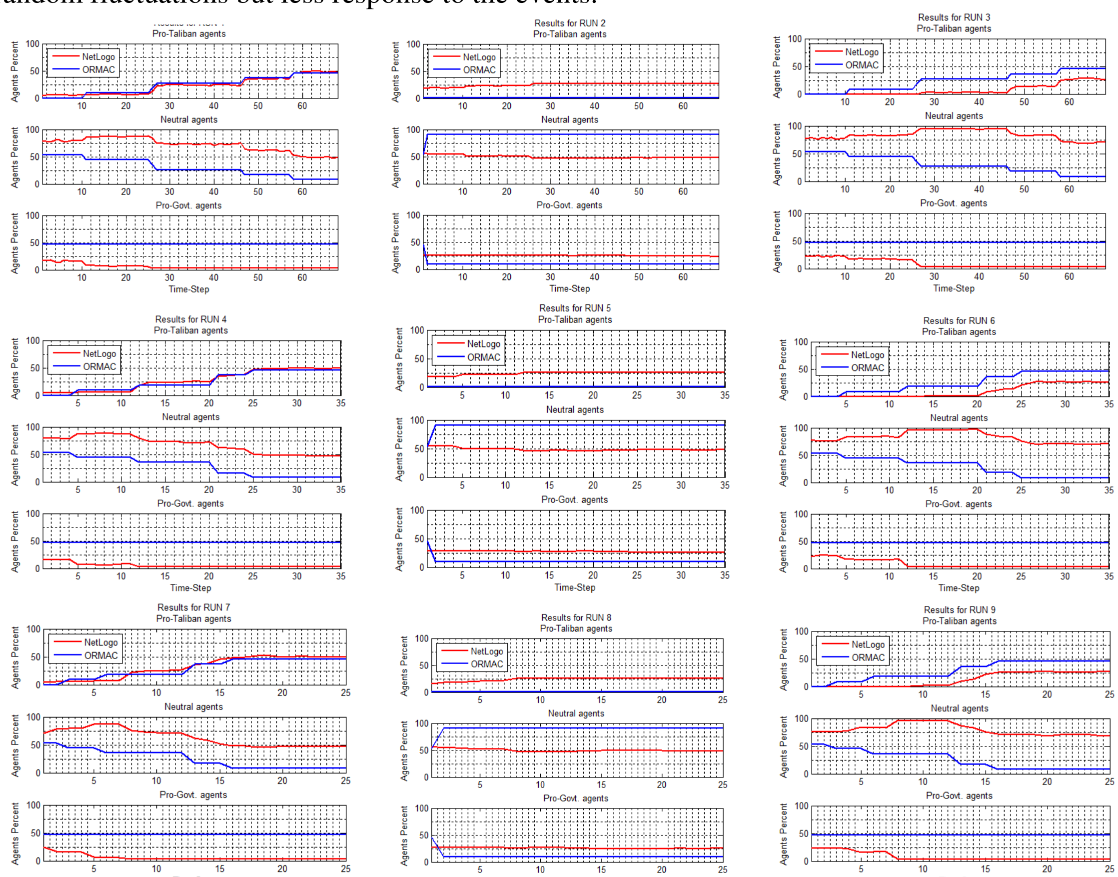


Figure 14: Nine simulation runs, as shown in Table 5, are shown in the nine panels, with the runs arranged sequentially from top to bottom. The outputs generated from ORMAC (blue) and NetLogo (red) correspond to the number of agents exhibiting pro-Taliban (top of each panel), neutral (middle) and pro-government (bottom) behavior.

The difference in the response to events versus random fluctuations is obvious from Fig. 15, which shows the first differences, or the differences in the outputs between successive time steps.

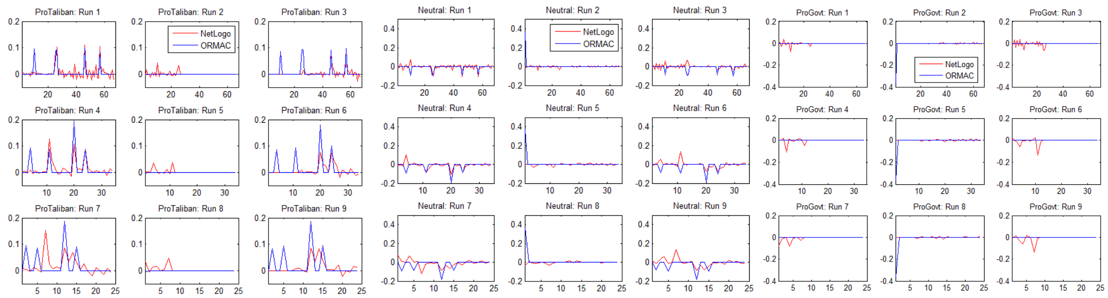


Figure 15: The first differences for the plots shown in Fig. 14, with each of the three panels (from left to right, with nine plots each) showing the nine runs in sequence (left to right then top to bottom). The three panels are, from left to right, for pro-Taliban, neutral, and pro-government classes, respectively, with NetLogo results in red and ORMAC in blue.

The first difference plots in Fig. 15 indicate that the response to the events is much clearer in ORMAC, and NetLogo generates more random fluctuations. Although a definitive explanation may not be possible without further investigation, the dominance of random parameters for initialization in NetLogo versus the more deterministic initializations and simulations in ORMAC may be a plausible explanation. In situations where both NetLogo and ORMAC exhibit what appears to be "legitimate" (i.e., occurring at the expected time steps) responses to events, the NetLogo responses seem somewhat damped compared to the ORMAC response. This indeed may be caused by resolution effects, even though there are a few exceptions to this empirical "rule." The lack of any ORMAC response (even the changes in NetLogo appear to be no more than random fluctuations) to the disruptive events when the representative (R) theory is instantiated may be worth noting. The lack of a leader may have less immediate effect on followers when the predominant behavior is representative. However such social explanations must be exercised with care given that the simulation runs appear pretty flat in each case when this theory is implemented (runs 2, 5, and 8). This may suggest an artifact of the specific experimental design. Although the causal explanations offered here are only plausible but not proven unless further simulations are performed, the value of simple metrics (e.g., bias and first differences) together with visual representations may be apparent from the discussions here. Traditional statistical distance measures are shown next. The correlation coefficients (CC) in Table 7 further indicate that the increase in the number of days per tick does not significantly improve the dependency between systems with respect to the number of agents in each class.

Table 7: Correlation coefficients between NetLogo and ORMAC outputs

Run	Days/tick	Theory	Pro-Taliban	Neutral	Pro-Govt
1	**3**	**L**	**0.9750**	**0.8870**	**-0.2867**
4	**7**	**L**	**0.9776**	**0.9463**	**-0.4136**
7	**14**	**L**	**0.9610**	**0.9109**	**-0.6806**
2	**3**	**R**	**-0.2505**	**-0.2614**	**0.1203**
5	**7**	**R**	**-0.3966**	**-0.4101**	**0.1839**
8	**14**	**R**	**-0.5601**	**-0.6048**	**0.2603**
3	**3**	**C**	**0.8479**	**0.1196**	**-0.2215**
6	**7**	**C**	**0.9361**	**0.4923**	**-0.3166**
9	**14**	**C**	**0.9479**	**0.5612**	**-0.4139**

If the pro-government behavior is ignored (because the ORMAC variations in that category are limited), the correlations are strongest for legitimacy (L) theory, followed by coercion (C). Correlations for representative (R) are negative, confirming the intuition that resolution effects may be very important for that theory. The mean squared errors are rather low (Table 8), suggesting that on average for all the simulation time steps, the errors are not too high.

Table 8: MSE between NetLogo and ORMAC outputs

Run	Days/tick	Theory	Pro-Taliban	Neutral	Pro-Govt
1	**3**	**L**	**0.0013**	**0.1804**	**0.1677**
2	**3**	**R**	**0.0644**	**0.1648**	**0.0244**
3	**3**	**C**	**0.0333**	**0.3132**	**0.1467**
4	**7**	**L**	**0.0018**	**0.1567**	**0.1715**
5	**7**	**R**	**0.0602**	**0.1697**	**0.0296**
6	**7**	**C**	**0.0284**	**0.3089**	**0.1532**
7	**14**	**L**	**0.0033**	**0.1521**	**0.1691**
8	**14**	**R**	**0.0598**	**0.1596**	**0.0263**
9	**14**	**C**	**0.0294**	**0.3159**	**0.1550**

A pair-wise comparison of the correlations between ORMAC and NetLogo runs based on instantiations of different leadership theories is shown in Tables 9 and 10. The results are interesting not just for exploring how the outputs from multiple theories relate to each other, but also for seeing the relations themselves change between the two implementations.

Table 9: Correlation coefficients between runs for NetLogo outputs

Runs	Days/tick	Theories	Pro-Taliban	Neutral	Pro-Govt
1,2	**3**	**L,R**	**0.7942**	**0.4615**	**0.6081**
1,3	**3**	**L,C**	**0.9166**	**0.4490**	**0.9096**
2,3	**3**	**R,C**	**0.5329**	**-0.4681**	**0.7442**
4,5	**7**	**L,R**	**0.7759**	**0.3778**	**0.5842**
4,6	**7**	**L,C**	**0.9149**	**0.5631**	**0.9198**
5,6	**7**	**R,C**	**0.5144**	**-0.4346**	**0.7077**
7,8	**14**	**L,R**	**0.8190**	**0.4502**	**0.6347**
7,9	**14**	**L,C**	**0.9378**	**0.6350**	**0.9006**
8,9	**14**	**R,C**	**0.6023**	**-0.2874**	**0.7761**

Table 10: Correlation coefficients between runs for ORMAC outputs

Runs	Days/tick	Theories	Pro-Taliban	Neutral	Pro-Govt
1,2	**3**	**L,R**	**-0.1879**	**-0.1879**	**-1.0000**
1,3	**3**	**L,C**	**0.9996**	**0.9996**	**1.0000**
2,3	**3**	**R,C**	**-0.1833**	**-0.1834**	**-1.0000**
4,5	**7**	**L,R**	**-0.2551**	**-0.2551**	**-1.0000**
4,6	**7**	**L,C**	**0.9996**	**0.9996**	**1.0000**
5,6	**7**	**R,C**	**-0.2503**	**-0.2505**	**-1.0000**
7,8	**14**	**L,R**	**-0.3626**	**-0.3627**	**-1.0000**
7,9	**14**	**L,C**	**0.9998**	**0.9998**	**1.0000**
8,9	**14**	**R,C**	**-0.3533**	**-0.3534**	**-1.0000**

The correlation between the outputs from the legitimacy (L) and coercion (C) theories, as implemented here, is high and appears to be least affected by the simulation granularity (i.e., NetLogo versus ORMAC). However, the relations appear completely altered for any comparison involving the representative (R) theory. Once again, this may be an artifact of the specific simulation, as all the plots (see Fig. 14) with this theory appear relatively flat. One other observation is that the change in number of days per tick seems to cause no statistical difference in the dependency between runs with respect to the number of agents in each class.

A detailed comparison of a relatively "lumped" or low-resolution model (e.g., the NetLogo-based implementation) with a relatively more spatially "distributed" or high-resolution model (e.g., the ORMAC-based implementation) typically entails one of two approaches: either aggregate the distributed model outputs to the scales of the lumped model and compare at the aggregate scales ,or allocate the lumped model outputs to the scales of the distributed model and compare at the higher resolutions. A comparison based on aggregation may be fairer because the allocation process is not necessarily well defined and may introduce errors. (In contrast, aggregation processes are typically well defined; for example when agent counts are considered, a simple sum would almost always be considered appropriate.) Indeed, most of the previous comparisons in this section would fall in that category. However, in situations where the end users or stakeholders demand higher resolution outputs, or when the underlying dominant social processes can be best captured at higher resolutions, a comparison at those resolutions may be more appropriate. In such cases, allocations become necessary for comparisons. In the absence of ancillary information, simple methods like equal or area-weighted allocations are probably the only options. In our case, the aggregate "patch" level outputs generated from NetLogo need to be allocated to the finer grids at which data are obtained from the Geographical Information Systems (GIS) and which are ultimately used by the ORMAC-based simulations. The simulation results must be compared at scales that matter to decision-makers (e.g., district levels in Afghanistan). The map for the case study region (Ghazni) with one NetLogo patch, corresponding ORMAC grids, and the Afghan districts (indicated by identification numbers assigned for the purpose of this simulation) is shown in Fig. 16.

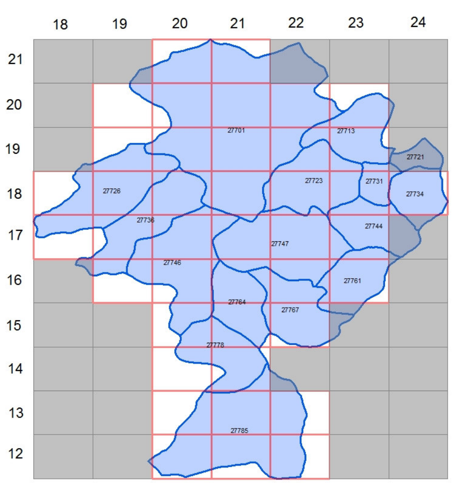


Figure 16: A map of the Ghazni region in Afghanistan that was used for the case study. The aggregate level NetLogo patch and the finer resolution ORMAC grids are indicated. The district boundaries are marked, and each district is assigned an identification number for the purposes of the simulations.

Motivated by the points discussed above, and driven by the need to develop proof-of-concept comparison metrics and methods, we uniformly allocated the aggregate NetLogo outputs to the resolutions of the ORMAC simulations. Specifically, the number of agents for each class and each patch was converted into the corresponding number of agents for each district by multiplying the uniformly distributed number of agents by the number of patches that equal the geographical size of each district. In a sense, this is just an area-weighted allocation strategy. We focused the comparison on "Run 1" (see Table 5) and a few districts for illustrative purposes.

Table 11: Linear correlation between ORMAC and NetLogo runs at district levels

RUN	District ID	Pro-Taliban	Neutral	Pro-Govt
1	27713	0.9744	0.8855	0.0000
	27723	0.9759	0.8890	0.0000
	27726	0.9759	0.8897	0.0000
	27731	0.9744	0.8856	-0.2867
	27747	0.9740	0.8846	0.0000

Table 11 shows high correlations between the ORMAC and NetLogo outputs for pro-Taliban and neutral agents. The zero or negative correlations for pro-government agents may be ignored given that the number of these agents remains relatively constant during the simulation time period (see Fig. 14, first panel, bottom plot). The high correlation at a district level indicates that a simple allocation was able to handle the resolution effect in terms of the correlation measure. However, these results should be generalized with caution and only if further experiments are confirmatory. In addition, we note that correlation is one measure that captures linear associations among data fluctuations, but more nuanced differences may not be obvious by looking at this measure alone. Thus, the effects of resolution are clearly seen in the district-level agent distributions over time, as shown in Fig. 17. In fact, Fig. 17 clearly shows that both the time at which predominant behavior changes (e.g., from dominant pro-Taliban to dominant neutral) and the magnitude of the change can be significantly impacted by the simulation and data resolutions (i.e., the ORMAC versus NetLogo simulations). Indeed, the final agent distributions are also significantly affected.

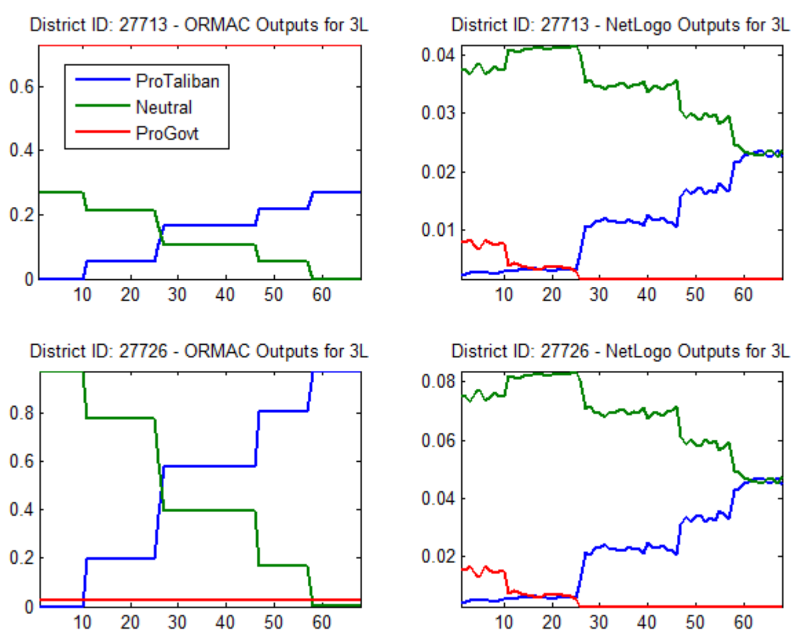


Figure 17: Distribution of agent fractions (the numbers are scaled) in two sample districts (see Table 11) from Run 1 (see Table 5). The effects of finer data and model resolutions (i.e., the differences between NetLogo and ORMAC outputs) are clearly seen in the overall magnitudes, the time when predominant behavior changes, and the magnitude of the change. The similarity to Fig. 5 is noted.

The similarity of Figures 17 and 5 is interesting from at least a couple of perspectives: (1) the possibility of interpreting Fig. 17 or similar plots in terms of "social emergence," with links to COA analysis, as in Fig. 5; (2) the possibility of obtaining equivalent representations of interpretable and useful emergence from disparate HSBC simulation strategies. For the purposes of this paper, we prefer to leave these as open areas for future research. However, we do point the reader to an intriguing, albeit unproven, possibility: The development of basic insights about social emergence from simple or assumed data, models and/or computational implementations, and the subsequent honing of these insights with enhanced data, models and computer power.

4.5 Distance Measures for Noisy or Limited Observations

Distance measures are routinely used to compare multiple simulations with each other and with observed data. However, the added complexity in HSBC M&S is that the models are more imperfect than usual, dependence across geographies and among entities may be manifest in different ways, and the effects of random noise or randomness in simulations as well as nonlinearities and thresholds may compound the difficulties associated with incomplete observations. In the previous section, we developed or utilized distance measures to compare the NetLogo- and ORMAC-based simulations. Distance measures may include bias and mean-squared errors, linear correlation, rank-based correlation measures like the Kendall’s Tau, and MI-based nonlinear correlations (Khan et al. 2007). In addition, quantile correlations (e.g., correlations among percentiles) and “threat scores” (Sabesan et al. 2007) have been utilized to explore correlations among extremes (Kuhn et al. 2007), or among low probability but high impact events. Despite considerable progress in the development of data resources to drive HSBC models, observations relevant for calibration and validation remain noisy and incomplete. Thus, the applicability of traditional statistical distance measures remains severely limited. The complexity of the problem is illustrated by the apparent lack of any linear association between related events (Fig. 15).

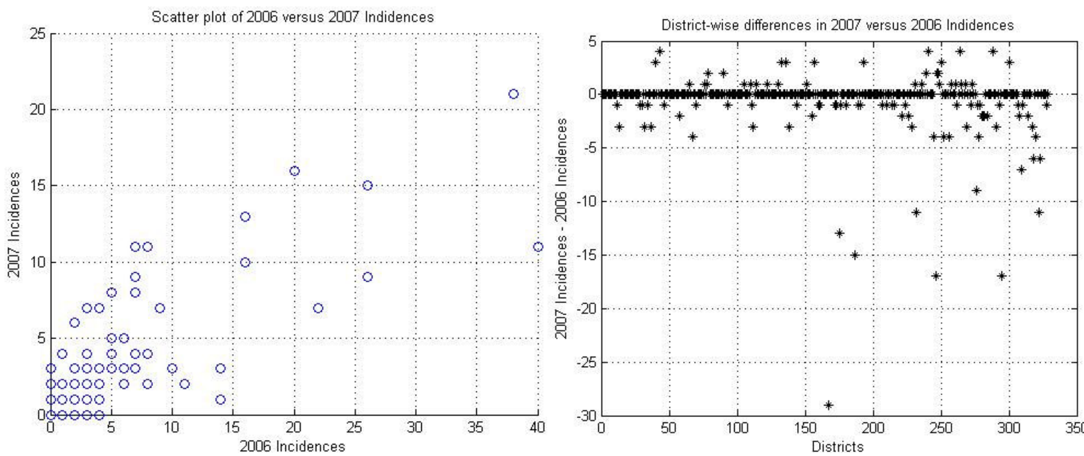


Figure 15: Scatter plot of 2006 versus 2007 terrorist incidences, and a district-wise difference plot of the year-to-year differences in the number of incidences, in a region of Afghanistan. The apparent lack of linear associations between the yearly data, as well as the lack of evident patterns in the errors, may be noted. The temporal resolution covering a full year, or the geospatial resolution spanning an entire district, may be grossly inadequate to calibrate models that may need to be run on timescales of days to weeks and at the granularity of individual social actors or organizations. However, such data availability situations are typical for HSBC calibration or validation.

5. Conclusions

This paper focuses on systematic evaluation of human, social, cultural, and behavioral (HSBC) modeling and simulation (M&S) systems. The experimental or simulation test-bed developed for the purpose (DARPA Foundation Team 2008), as well as the limitations of any real-world insights drawn from the test-bed at this stage, has been described earlier. The systematic evaluation tasks inherit many of those limitations. On the whole, we urge caution before generalizing the insights or conclusions developed in this article to the real world. However, we believe that an important and promising step, albeit small, has been taken toward achieving the ultimate goal.

The purpose of this paper was to demonstrate, in a preliminary and proof-of-concept fashion, the feasibility of the following in the context of social science simulations:

1. Perform structural and parametric sensitivity analysis for causal analysis and detection or prediction of socially relevant emergence, with the ultimate goal of developing best-fit model and theory selection and recommendation strategies.
2. Characterize emergence types and develop ways to quantify social emergence, as well as metrics to characterize HSBC M&S systems that are capable of generating emergence in terms of microscale evolution or rules or dynamical processes and in terms of macroscale signatures. (We acknowledge that no one definition of emergence is available or prevalent, and perhaps such a universal definition is neither necessary nor desirable. However, operational definitions of emergence from multiple considerations remain useful.)
3. Develop trade-offs between emergence and predictability in HSBC M&S systems with the purpose of characterizing such systems and generating recommendations for use based on end user requirements.
4. Develop methods to extract dominant social processes from observations, simulation outputs, and human insights, with the purpose of enhancing understanding of the underlying social phenomena, suggesting best models or model combinations depending on which processes dominate, and producing insights that can be used by military operators (commanders and strategists).
5. Develop distance measures to compare multiple simulation results, as well as to compare simulations with observations (even when such observations are noisy, sparse, partial, or incomplete), with the goal of evaluating performance of HSBC systems in terms of modeling predominant behavior and processes, extreme values, nonlinear and rare processes, tipping points, and surprising behavior.

The area of HSBC M&S suffers from models that are poorly understood (relative to models for most physical, built, or natural systems) and data that are inherently noisy, sparse, and incomplete. Thus, validation takes on the form of characterization and systematic evaluation, with the ultimate aim of providing value to end users and stakeholders, in this case military commanders. The results presented here take a first step in this challenging direction.

References

1. Ajzen, I. (2001): *Nature & operation of attitudes*, **Annual Review of Psychology**, 52:27–58.
2. Alessa, L.N., Laituri, M., M. Barton (2006): *An "All Hands" Call to the Social Science Community: Establishing a Community Framework for Complexity Modeling Using Agent Based Models & Cyberinfrastructure*, **J. of Artificial Societies & Social Simulation**, 9(4).
3. Becker, G. (1968): *Crime and punishment*, **Journal of Political Economy**, 76:196–217.
4. Berger, J., Ridgeway, C.L., Fisek, M.H., and R.Z. Norman (1998): *The legitimation and delegitimation of power and prestige orders*, **American Sociological Review**, 63:379–405.
5. Beven, K., and A.M. Binley (1992): *The future of distributed models: model calibration and uncertainty prediction*, **Hydrological Processes**, 6:279–298.
6. Bhaduri, B., Bright, E., Coleman, P., and J. Dobson (2002): *LandScan: locating people is what matters*, **Geoinformatics**, 5(2):34–37.
7. Bhattacharjee, Y. (2007): *Pentagon asks academics for help in understanding its enemies*, **Science**, 316(5824):534–535.
8. Bonabeau, E. (2002): *Agent-based modeling: methods and techniques for simulating human systems*, **Proc. of the National Academy of Sciences**, 99(10), Supplement 3, 7280–87.
9. Bourdieu, P. (1984): **Distinction: A Social Critique of the Judgment of Taste**, Cambridge, Massachusetts: Harvard University Press.
10. Boschetti, F., Prokopenko, M., Macreadie, I., and A.-M. Grisogno (2005): *Defining and detecting emergence in complex networks*, In Khosla, R., et al. (Eds.), **Knowledge-Based Intelligent Information and Engineering Systems,** Lecture Notes in Computer Science 3684, pp. 573–580, Springer-Verlag.
11. Box, G., Jenkins, G.M., and G. Reinsel (1994): **Time Series Analysis: Forecasting and Control**, 3rd Edition, Prentice Hall.
12. Brown, D.G., Riolo, R., Robinson, D.T., North, M., and W. Rand (2005): *Spatial process and data models: toward integration of agent-based models and GIS*, **Journal of Geographical Systems**, 7(1):25–47.
13. Bryson, J.J., Ando, Y., and H. Lehmann (2007): *Agent-based modelling as scientific method: a case study analysing primate social behaviour*, **Philosophical Transactions of the Royal Society, B -- Biology**, 362(1485):1685–1698.
14. Calhoun, C. (1994): **Neither Gods Nor Emperors: Students and the Struggle for Democracy in China**, Berkeley: University of California Press.
15. Carley, K. (1986): *An approach for relating social structure to cognitive structure*, **Journal of Mathematical Sociology**, 12:137–89.
16. Carpenter, T.M., and K.P. Georgakakos (2006): *Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales*, **Journal of Hydrology**, 329(1–2):174–185.
17. Chapman, J., and P. Winker (2005): *Optimal aggregation of linear time series models*, **Computational Statistics & Data Analysis**, 49(2):311–331.
18. Chaturvedi, A.R., Dolk, D., Chaturvedi, R., Malpuri, M., Lengacher, D., Mellema, S., Poddar, P., Foong, C., and B. Armstrong (2005): *Understanding insurgency by using agent-based computational experimentation: case study of Indonesia*, **Proceedings of the Agent 2005 Conference on Generative Social Processes, Models, and Mechanisms**, ANL/DIS-06-5, ISBN 0-9679168-6-0, C.M. Macal, M.J. North, and D. Sallach (Eds), co-sponsored by Argonne National Laboratory and The University of Chicago, October 13−15.
19. Coleman, J.S. (1990): **Foundations of Social Theory**, Cambridge, Massachusetts: Harvard University Press.
20. Collier, N.T. (2003): **RePast: An Extensible Framework for Agent Simulation**, Social Science Research Computing, University of Chicago, 18 pages.

21. Cover, T.M., and J.A. Thomas (2006): **Elements of Information Theory** (2nd ed.). New Jersey: Wiley Interscience.
22. Cressie, N.A. (1993): **Statistics for Spatial Data**, Wiley Series in Probability and Statistics.
23. DARPA Foundation Team (2008): **Actionable Capability for Social and Economic Systems (ACSES)**, Seedling Report submitted to DARPA.
24. Crutchfield, J.P. (1994): *The Calculi of Emergence: Computation, Dynamics, and Induction*, **Physica D**, 75 (1-3): 11-54.
25. Deguet, J., Demazeau, Y., and L. Magnin (2006): *Elements about the emergence issue: a survey of emergence definitions*, **Complexus**, 3:24–31.
26. de Haan, J. (2006): *How emergence arises*, **Ecological Complexity**, 3:293–301.
27. Dessalles J.L., Ferber J., and D. Phan (2007): *Emergence in agent based computational social science: conceptual, formal and diagrammatic analysis*, in Shan Y. and Yang A. (Eds.), **Intelligent Complex Adaptive Systems**, Idea Group Inc, forthcoming.
28. Dobson, J.E., Bright, E.A., Coleman, P.R., Durfee, R.C., and B.A. Worley (2000): *LandScan: a global population database for estimating populations at risk*, **Photogrammetric Engineering and Remote Sensing**, 66(7):849–857.
29. Done, J.M., Davis, C.A., Leung, L.R., and B. Kuo (2004): *Understanding the value of high resolution regional climate modeling*, **Eos Transactions of the American Geophysical Union**, 85(47), Fall Meeting Supplement, Abstract 32A-11.
30. Doubilet, P., Begg, C.B., Weinstein, M.C., Braun, P., and B.J. McNeil (1985): *Probabilistic sensitivity analysis using Monte Carlo simulation*, **Medical Decision Making**, 5(2):157–177.
31. Epstein, J.M. (2007): **Generative Social Science: Studies in Agent-Based Computational Modeling**, Princeton University Press, 352 pages.
32. Evans (2005): *Deductive reasoning*, in Holyoak, K.L. and R.G. Morrison (eds.), **The Cambridge Handbook of Thinking and Reasoning**, 169-184.
33. Fadlalla, A. (2005): *An experimental investigation of the impact of aggregation on the performance of data mining with logistic regression*, **Information & Management**, 42(5):695–707.
34. Fagiolo, G., Birchenhall, C. and P. Windrum (Eds.) (2007): *Special issue on empirical validation in agent-based models*, **Computational Economics**, 30(3).
35. Festinger, L. (1957): **A Theory of Cognitive Dissonance**, Stanford, California: Stanford University Press.
36. Fukunaga, K. (1990): **Introduction to Statistical Pattern Recognition**, 2nd Edition, Academic Press, 592 pages.
37. Gamerman, D., and H.F. Lopes (2006): **Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference**, CRC Press, 323 pages.
38. Ganguly, A. R., and R.L. Bras (2003): *Distributed quantitative precipitation forecasting using information from radar and numerical weather prediction models*, **Journal of Hydrometeorology**, 4:1168–1180.
39. Goertzel, B. (1993): **The Structure of Intelligence: A New Mathematical Model of Mind**, Springer, 187 pages.
40. Grassberger, P., and I. Procaccia (1983): *Characterization of strange attractors*, **Physical Review Letters**, 50(5):346–349.
41. Grebogi, C., Ott, E., and J.A. Yorke (1987): *Chaos, strange attractors, and fractal basin boundaries in nonlinear dynamics*, **Science**, 238(4827):632–638.
42. Han, J., and M. Kamber (2006): **Data Mining: Concepts and Techniques**, Elsevier, 800 pp.
43. Haykin, S. (1994): **Neural Networks: A Comprehensive Foundation**, 1st Edition, Prentice Hall, 768 pages.
44. Hazen, G.B., and M. Huang (2006): *Parametric sensitivity analysis using large-sample approximate Bayesian posterior distributions*, **Decision Analysis**, 3(4):208–219.
45. Higham, D.J., and N.J. Higham (2005): **MATLAB Guide**, SIAM, 406 pages.

46. Houtekamer, P., and H.L. Mitchell (1998): *Data assimilation using an ensemble Kalman filter technique*, **Monthly Weather Review**, 126(3):796–811.
47. Hudson, V.M., Schrodt, P.A., and R.D. Whitmer (2008): *Discrete sequence rule models as a social science methodology: an exploratory analysis of foreign policy rule enactment within Palestinian-Israeli event data*, **Foreign Policy Analysis** (forthcoming).
48. Hughes, B.B. (1999): *The International Futures (IFs) Modeling Project*, **Simulation & Gaming**, 30(3):304–326.
49. Husslage, B., Rennen, G., van Dam E.R., and D. den Hertog (2006): *Space-filling Latin hypercube designs for computer experiments*, Working Paper No. 2006-18, ISSN 0924-7815, Department of Econometrics and Operations Research, Tilburg University, March 2006.
50. Jordan, M.I., and R.A. Jacobs (2001): *Hierarchical mixture of experts and the EM algorithm*, In M.I. Jordan and T.J. Sejnowski (Eds), **Graphical Models: Foundations of Neural Computation**, chapter 11, pp. 257–290.
51. Judd, K., and L.A. Smith (2004): *Indistinguishable states II: the imperfect model scenario*, **Physica D**, 196:224–242.
52. Khan, S., Bandyopadhyay, S., Ganguly, A.R., Saigal, S., Erickson III, D.J., Protopopescu, V., and G. Ostrouchov (2007): *Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data*, **Physical Review E**, 026209.
53. Khan, S., Ganguly, A.R., and S. Saigal (2005): *Detection and predictive modeling of chaos in finite hydrological time series*, **Nonlinear Processes in Geophysics**, 12:41–53.
54. Kiser, E., and A. Linton (2002): *The hinges of history: state-making and revolt in early modern France*, **American Sociological Review**, 67:889–910.
55. Kleijnen, J.P.C. (1998): *Experimental design for sensitivity analysis, optimization and validation of simulation models*, In J. Banks (Ed.), **Handbook of Simulations: Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice**, 6:173–224.
56. Krishnamurti, T.N., Kishtawal, C.M., LaRow, T.E., Bachiochi, D.R., Zhang, Z., Williford, C.E., Gadgil, S., and S. Surendran (1999): *Improved weather and seasonal climate forecasts from multimodel superensemble*, **Science**, 285(5433):1548–1550.
57. Kuhn, G., Khan, S., Ganguly, A.R., and M. Branstetter (2007): *Geospatial-temporal dependence among weekly precipitation extremes with applications to observations and climate model simulations in S. America*, **Advances in Water Resources**, 30(12):2401-2423.
58. Kuipers, B. (1994): **Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge**, MIT Press, 452 pages.
59. Langton, Chris G. (1991): *Computation at the edge of chaos: phase transitions and emergent computation*. In Stephanie Forrest (Ed.), **Emergent Computation**. Cambridge, Massachusetts: The MIT Press.
60. Lee, W., and D. Xiang, (2001): Information-theoretic measures for anomaly detection. *Proceedings of the IEEE Symposium on Security and Privacy*, 130–143.
61. Legasto, A., Forrester, J.W., and J.M. Lyneis (1980): **System Dynamics**, Elsevier, 282 pages.
62. Levi, M. (1988): **Of Rule and Revenue**. Berkeley: University of California Press.
63. Lewis-Beck, M.S., Bryman, A. and T.F. Liao (2004): **The Sage Encyclopedia of Social Science Research Methods**, Sage Publications.
64. Ma, J., James, T., and P. Simon (2003): *Accurate on-line support vector regression*. **Neural Computation**, **15**:2683–2703.
65. Machiavelli, N. 1985 [1513]: **The Prince**. Chicago: University of Chicago Press.
66. MacKay, D. (1995): *Probable networks and plausible predictions—a review of practical Bayesian methods for supervised neural networks*, **Network: Computation in Neural Systems**, 6:469–505.
67. Magnani, L., Narsessian, N.J., and P. Thagard (1999): **Model-Based Reasoning in Scientific Discovery**, Springer, 360 pages.

68. Marks, R. (2008): *Validating simulation models: a general framework and four applied examples*, **Computational Economics**, 30(3):265–290.
69. Marwell, G., and P. Oliver (1993): **The Critical Mass in Collective Action. A Micro-Social Theory**, New York: Cambridge University Press.
70. Marx, K. 1990 [1867]: **Capital**, Penguin.
71. McPhail, C. (1991): **The Myth of the Madding Crowd**, New York: Aldine De Gruyter.
72. McPhail, C. (1994): *The dark side of purpose: individual and collective violence in riots*, **Sociological Quarterly**, 35:16–32.
73. Mills, T. (1991): **Time Series Techniques for Economists**, Cambridge Univ. Press, 387 pp.
74. Miller, J.H. (1998): *Active nonlinear tests (ANTs) of complex simulation models*, **Management Science**, 44(6):820–830.
75. Murphy, A.H. (1993): *What is a good forecast? An essay on the nature of goodness in weather forecasting*, **Weather Forecasting**, 8:281–293.
76. Myers, D.J., and P.E. Oliver (2000): *The opposing forces diffusion model: the initiation and repression of collective violence*, Working Paper, The University of Notre Dame.
77. Nisbett, R.E., and L. Ross (1980): **Human Inference: Strategies and Shortcomings of Social Judgment**, Englewood Cliffs, New Jersey: Prentice-Hall.
78. Ogata, K. (1998): **System Dynamics**. Prentice Hall, 758 pages.
79. Oliver, P. (1993): *Formal models of collective action*, **Annual Review of Sociology**, 19:271–300.
80. Olson, J.M., Vernon, P.A., Harris, J.A., and K. Jang (2001): *The heritability of attitudes: a study of twins*, **Journal of Personality and Social Psychology**, 80:845–860.
81. Opp, K.-D., and C. Gern (1993): *Dissident groups, personal networks, and spontaneous cooperation: the East German revolution of 1989*, **American Sociological Review**, 58:659–680.
82. Opp, K.-D., and W. Roehl (1990): *Repression, micromobilization, and political protest*, **Social Forces**, 69:521–548.
83. Ostrom, E. (2007): *Collective Action Theory*, pp. 186–208 in **The Oxford Handbook of Comparative Politics**, Boix, C., Stokes, S. C., (Eds), Oxford: Oxford University Press.
84. Parenti, C. (2007): *Taliban rising*, **The Nation**, Posted online at http://www.thenation.com/doc/20061030/parenti_video on 10/12/2007.
85. Petty, R.E., Wegener, T.T., and L.R. Fabrigar (1997): *Attitudes and attitude change*, **Annual Review of Psychology**, 48:609–647.
86. Potok, T., Elmore, M., Reed, J., and F.T. Sheldon (2003): *VIPAR: advanced iInformation agents discovering knowledge in an open and changing environment*, **Proceedings of the 7th World Multiconference on Systemics, Cybernetics and Informatics: Special Session on Agent-Based Computing**, Orlando, Florida, pp. 28–33, July 27–30.
87. Potok, T.E., Ivezic, N.A, and N.F. Samatova (2000): *Agent-based architecture for flexible lean cell design, analysis, and evaluation*, **Proceedings of the 4th Design of Information Infrastructure Systems for Manufacturing Conference**.
88. Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., and DMIP Participants (2004): *Overall distributed model intercomparison project results*, **Journal of Hydrology**, 298(1–4):27–60.
89. Rabiner, L.R. (1989): *A tutorial on hidden Markov models and selected applications in speech recognition*, **Proceedings of the IEEE**, 77(2):257–286.
90. Railsback, S.F., Lytinen, S.L., and S.K. Jackson (2006): *Agent-based Simulation Platforms: Review and Development Recommendations*, **Simulation**, 82 (9): 609-623.
91. Reynolds, C.W. (1987): *Flocks, herds and schools: a distributed behavioral model*, **ACM SIGGRAPH Computer Graphics**, 21(4):25–34.
92. Roscigno, V. J., and W.F. Danaher (2004): **The Voice of Southern Labor: Radio, Music, and Textile Strikes, 1929–1934**, Minneapolis: UMN Press.

93. Sabesan, A., Abercrombie, K., Ganguly, A.R., Bhaduri, B.L., Bright, E.A., and P. Coleman (2007): *Metrics for the comparative analysis of geospatial datasets with applications to high resolution grid-based population data*, **GeoJournal**, 30:2401–2423.
94. Saltelli, A. (2004): **Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models**, John Wiley and Sons, 232 pages.
95. Santner, T.J., Williams, B.J., and W. I. Notz (2003): **The Design and Analysis of Computer Experiments**, Springer Series in Statistics, 283 pages.
96. Sidman, Murray (2000): **Coercion and Its Fallout**, Boston, Massachusetts: Author's Cooperative.
97. Silverman, B.G., Johns, M., Cornwell, J., and K. O'Brien (2006): *Human behavior models for agents in simulators and games: part I: enabling science with PMFserv*, **Presence: Teleoperators and Virtual Environments**, 15(2):139–162.
98. Sterman, J.D. (2000): **Business Dynamics: Systems Thinking and Modeling for a Complex World**, Irwin / McGraw-Hill.
99. Sterman, J., and H. Rahmandad (2008): *Heterogeneity and network structure in the dynamics of diffusion: comparing agent-based and differential equation models*, **Management Science** (forthcoming).
100. Stigler, G. (1970): *The optimum enforcement of laws*, **Journal of Political Economy**, 78:526–536.
101. Subrahmanian, V.S. (2007): *Computer science: cultural modeling in real time*, **Science**, 317(5844):1509–1510.
102. Sugihara, G., and R.M. May (1990): *Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series*, **Nature**, 343(6268):734–741.
103. Tenenbaum, J.B., Griffiths, T.L. and C. Kemp (2006): *Theory-based Bayesian models of inductive learning and reasoning*, **Trends in Cognitive Science**, 10(7):309-318.
104. Tesfatsion, L., and K.L. Judd (Eds) (2006): **Handbook of Computational Economics, Vol. 2: Agent-Based Computational Economics**, Handbooks in Economics Series, North-Holland/Elsevier, Amsterdam.
105. Tisue, S., and U. Wilensky (2004): *NetLogo: A simple environment for modeling complexity*, **International Conf. Complex Systems**, Boston, Massachusetts, May 16–21.
106. Tversky, A., and D. Kahneman (1981): *The framing of decisions and the psychology of choice*, **Science**, 211:453–458.
107. Vicsek, T. (2002): *Complexity: the bigger picture*, **Nature**, 418:131.
108. Walker, I., and H.J. Smith (2001): **Relative Deprivation: Specification, Development, and Integration**, Cambridge University Press, 379 pages.
109. Weber, M. 1968 [1922]: **Economy and Society**, Totowa, New Jersey: Bedminster Press.
110. Weigend, A.S., and N.A. Gershenfeld (1994): **Time Series Prediction: Forecasting the Future and Understanding the Past**, Westview Press.
111. Whitmeyer, J.M. (2002): *Elites and popular nationalism*, **British Journal of Sociology**, 53(3):321–341.
112. Wickham-Crowley, T.P. (1993): **Guerrillas and Revolution in Latin America: A Comparative Study of Insurgents and Regimes since 1956**, Princeton Univ. Press.
113. Windrum, P., Fagiolo, G., and A. Moneta (2007): *Empirical validation of agent-based models: alternatives and prospects*, **J. of Artificial Societies and Social Simulation**, 10(2,8).
114. Wojna (2005): *Analogy-based reasoning in classifier construction*, **Transactions on Rough Sets IV**, Lecture Notes in Computer Science, 3700:277-374.
115. Wuensche, A. (1999): *Classifying cellular automata automatically: Finding gliders, filtering, and relating space-time patterns, attractor basins, and the Z parameter*, **Complexity**, 4(3):47–66.
116. Zeigler, B.P., Kim, T.G., and H. Praehofer (2000): **Theory of Modeling and Simulation**, 2nd Edition, Academic Press, 510 pages.