

StateSim: Lessons Learned from 20 Years of A Country Modeling and Simulation Toolset*

By

Barry G. Silverman, PhD^{1,2}, Daniel M. Silverman, PhD^{2,3}, Gnana Bharathy, PhD^{2,4}, Nathan Weyer², and William Tam²

1 – Electrical and Systems Engineering Dept, University of Pennsylvania, Philadelphia, PA 19104, barryg@upenn.edu

2 – Edutainiacs.com, LLC, Narberth, PA 19072, service@edutainiacs.com

3 – Institute for Politics and Strategy, Carnegie Mellon University, Pittsburgh, PA 1521

4 – School of Information, Systems & Modeling, University of Technology Sydney, Sydney, NSW 2007, AU

November 2020

- *The crucial issue is no longer "What do we know?" but rather*
- *"How do we deal with the fact that we don't know enough?"*
- - Attributed to C. West Churchman
-

ABSTRACT

A holy grail for military, diplomatic, and intelligence analysis is a valid set of software agent models that act as the desired ethno-political factions so that one can test the effects of alternative courses of action in different countries. This article explains StateSim, a country modeling approach that synthesizes best-of-breed theories from across the social sciences and that has helped numerous organizations over 20 years to study insurgents, gray zone actors, and other societal instabilities. The country modeling literature is summarized (Sect 1.1) and synthetic inquiry is contrasted with scientific inquiry (Sect. 1.2 and 2). Section 2 also explains many fielded StateSim applications and 100s of past acceptability tests and validity assessments. Section 3 then describes how users now construct and run ‘first pass’ country models within hours due to the StateSim Generator, while Section 4 offers two country analyses that illustrate this approach. The conclusions explain lessons learned.

Keywords: country modeling, sociological game theory, systems approach, cognitive agents, policy analysis tools

* - This article was the keynote talk of the 2020 Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation (SBP-BRiMS) Annual Conference. This article also outlines the program of research and development for which the 1st author was awarded the 2020 SBP-BRiMS Lifetime Achievement Award.

1) Introduction

Interstate and intrastate instabilities are predicted to increase in frequency and scale, particularly for weakly institutionalized low- and middle-income states being faced by various insurgencies, (Gray Zone) adversaries, and/or economic and climate stressors. The Gray Zone is the space between peace and outright war where an adversary is seeking to destabilize a regime, often from within, in order to have indigenous proxies and/or insurgents take over. It is vital to model and simulate how to influence these various instabilities in order to support societies facing these threats and help ensure peace and security around the globe.

StateSim is a toolset for modeling states, sub-states and sets of states. StateSim seeks to recreate a given ethno-political situation and community in an artificial society framework. Doing so permits user-analysts to explore what issues are driving a given community and what might influence it. Specifically, StateSim provides users with the ability to rapidly (within hours) model countries and/or regions so that they can study and explore the impact of Diplomatic, Informational, Military, and Economic (DIME) actions they might undertake to try and either enhance or challenge a given regional group's stability.

This article reviews the almost 20-year history of building, testing, validating, and fielding a continuing series of improved versions of StateSim. StateSim follows the systems approach and is a model of many sub-models. The social sciences are often reductive and good scientists are rewarded (in terms of tenure, promotions, publications, etc) for rigorously following the analytical scientific method. As Kuhn (1970) points out, disciplinary silos are paradigmatic and constrain science to work analytically to deepen the silo before advancing to the next one. This results in the decomposition and narrowing of topics – different college departments and different journals for different aspects of social science like politics, sociology, economics, psychology, etc. By contrast, StateSim has always been, and continues to be, a complexity science experiment in the synthetic paradigm, of crossing scientific barriers, of unifying best of breed theories across fields. It can only predict the primary, secondary, and tertiary effects of actions and operations if the agents in the game can reason across a context spanning the disciplines.

The rest of Section 1 surveys the literature on alternative country modeling approaches and where StateSim fits within that taxonomy (Sect 1.1). Section 1.2 delves deeper into the synthetic paradigm. StateSim models have successfully passed numerous validity assessments and acceptability tests over the years, many of which we review in Section 2. This section also traces the evolution of the components and sub-models that were tested, improvements made, and versions or applications fielded. In early versions of StateSim, profiling the countries was arduous. Section 3 reviews research on and changes made across the years that have eased the creation of country models. It also provides data on the duration of various model building exercises revealing two orders of magnitude of speedup (from weeks to days to hours). Users want speedup, but they also want to be sure StateSim allows them to explore courses of action and policy interventions across the DIME spectrum. To facilitate this, Section 4 provides two case study societies and how the agents behave in the base case as well as under various policy interventions. These are real-world scenarios and they allow the reader to judge their realism directly. Finally, Section 5 includes conclusions and lessons learned.

1.1) State-of-the Art For Country Modeling

According to Elsaesser et al (2015) the mathematical computation approaches to forecasting national level events tend to be either statistical forecasting models or complexity science based simulations. This section compares these two major approaches to country modeling. We conclude with an agent based approach (cognitive agents) that complements both of the other approaches and that serves to implement a sociological game theory (SGT).

Big Data and Black Box Statistical Modeling - The state-of-the-art of predictive modeling of nations and networks is largely theory-free model fitting to 'big data' that points inductively towards factors that have relatively strong predictive power. Here several researchers have been investigating "where" and "when" will the violence occur; and under what local conditions will outgroups become more organized and, thereby, more threatening. For instance, Goldstone, Bates, et al (2010) built a small regression model successfully able to forecast violent and non-violent regime reversals within a 2 year window (over 80% accuracy). According to Halkia et al (2020) the largest social event datasets are (i) the Global Data on Events Location and Tone (GDELT) project and (ii) the Integrated Crisis Early Warning System (ICEWS) Dataverse. They also state that due to noise (duplicate events), few have used GDELT for predictions. The ICEWS Dataverse dataset (actors, targets, event types) supports tracking and forecasting of sociopolitical conflicts and is available on the Harvard Dataverse Network: Boschee et al (2018), Jontz (2015). The ICEWS program itself uses the data to forecast disruptive events for 177 countries - from coups to rebellions, terrorist attacks, political crises and civil strife. ICEWS uses an ensemble of statistical models to make its forecasts (regressions, neural nets, Bayesian classifiers). Many of the forecasts were shown to have well over 90% accuracies during DARPA trials. There is little doubt that this sort of exercise has led to models that are useful as early warning systems for decision makers. At the same time, this class of models are theory-free black boxes and cannot be inspected

for causes of the instabilities. Nor can they be used in readiness training or analytic scenario analysis exercises to explore how to handle the conflict event or instability they are predicting. This is the reason for turning to the complexity science based simulations.

Pucksters and Branching Script Agents -- The largest sponsor of readiness and analytic simulation studies is the US military. The overwhelming state-of-the-practice in that domain is the use of human pucksters controlling avatars in various simulated spaces and following scripts to play out the roles of the diverse actors in the country of interest. Examples of such simulated spaces are OneSAF, Real World, Unreal Tournament and VBS3, among others. Using pucksters can be expensive, especially if there is a lot of readiness training and/or plan analyses to carry out in short order. In these worlds some of the low level actor functions are delegated by the pucksters to simple branching script or Finite State Machine (FSM) agents. These agents carry out basic activities such as physical interaction and point A to B movement in the terrain, tactical battle kinetics (eg, fire if fired upon), and carrying out routine daily activities to provide users with greater sense of immersion and place.

Perhaps the pinnacle of this scripted agent approach is the US Army's (and JPL developed) Athena simulation software: (Cape & Lee, 2019, NASA, 2011). Athena supports an analyst in making 2 to 3 year stability and reconstruction projections for operational analyses in regions and countries like Jordan, Iraq, Afghanistan, and others. It lets users explore diplomatic, information, military, and economic (DIME) actions and plots their effects on political, military, economic, social, information, and infrastructure (PMESII) aspects of society. This lets trainees and analysts explore how DIME interventions cause changes in neighborhood security, leader and group relationships, population attitudes, and quality of life.

Overall, the agents in these systems are scripted, simplistic and rigid, though computationally efficient and, hence, highly scalable in terms of numbers of agents. When simulation scenarios unfold according to script, these agents can be a very effective way to enhance the player's immersion and sense of place. However, simulation exercises often lead to unexpected situations or branches that the scripts don't cover and thereby the agents weren't pre-programmed to handle. The problem is that once simulation begins, the original assumptions behind the pre-composed behaviors might no longer hold, and scripted agents will not have any way to reason about a new situation.

Cognitive Agents As a Complement/Extension – The 'big data' forecasting approach is best at predicting country events and trends and the scripted agent simulations are scalable for readiness training and predictable plan analysis. They both are successful at what they do since they use simple models – the KISS approach. However, one results in a black box that can't be probed for what-ifs and the other can only handle what-ifs that are pre-scripted. The cognitive agent approach of StateSim has often been used to complement each of these approaches as we will explain in this article.

If agents are to reason better about the scenarios that unfold around them, some investigators argue they need to be game theoretic and aware of the games that other agents are playing and subjecting them to. Analytic game theory is the mathematics of strategy, and as such, holds great promise for helping to understand conflicts. At the same time, analytic game theory has a weak record of explaining and/or predicting real world conflict – about the same as random chance according to Armstrong (2002).

Real world agents are rational (maximizing their own goals), but only if you understand their moralistic value system as well as the socio-economic-political-religious-ethnic contexts they are responding to. In the field of social science, Burns et al (2017) point out that the explanatory and predictive powers of analytic game theory can be improved by replacing prescriptions from rational economics with descriptions from sociology (what Edmonds & Moss (2015) call Keep It Descriptive, Stupid, KIDS). This is also articulated in Burns et al (2017)'s 'sociological game theory' which adds in "norm, value, belief, role, social relationship, and institution". As per Camerer and Ho (2015), we seek to include emotions and bias into subjective utility processes, something that Askari et al (2019) call agent hyper-rationality. This integration of norms, emotions, and biases with traditional ideas of material self-interest has been successfully incorporated into research programs on issues as diverse as the functioning of economic markets (Benabou 2013) and the information landscapes of war zones in recent years (Silverman 2018). StateSim pursues such a descriptive profiling approach. As we shall demonstrate, this approach leads to accurate recreations and predictions of the agents, factions, and institutions being profiled. And since it uses (white or gray box) causal modeling, this permits re-running scenarios and exploring the impacts of what-if courses of action and counterfactuals.

1.2) The Synthetic Paradigm

The intellectual roots of StateSim are illustrated by a request we responded to in 2000 from the Defense Modeling and Simulation Office. They were concerned that the field of human behavior modeling was too conservative and investigators tended to work in silos and ignore each other's results. For example, there were individual dose-response models of how human performance was affected by various moderators like battlefield noise, inter-soldier loyalty, or injury, but no cross-cutting or unifying models. As another example, Janis & Mann (1977) stated that in order to model

human stress one needs to combine emotion models from psychology, fatigue models from medicine, and time pressure models from management. Such a model has still yet to be assembled and scientifically tested, although all the component models have been. There are innumerable other examples like that. We illustrate a final example in Section 3.1 and 3.2 of how people strategize in conflict vs cooperation. There we demonstrate the need to synthesize game theory, behavioral game theory, the prevailing psychological theory of decision making (ie, cognitive appraisal), and personality and ethnic profiling instruments. Without the combination of all these diverse elements, no one of them alone can achieve a high degree of validity, as we will demonstrate.

The HBM field and the computational social sciences are sorely missing the synthetic paradigm that reaches across disciplines and sweeps in best of breed theories. More researchers must use the synthetic paradigm and conduct experiments that help the field learn how best to integrate and fuse the many relevant components that have been published. This is a suggestion for embracing the synthetic paradigm, which is the core of systems science. For that purpose, an important branch of the systems field was Pragmatism, the form of synthetic inquiry founded by the philosopher EA Singer, Jr (19xx). He suggested “sweeping in” the “outside” viewpoints of other fields that may be thought of as the enemies of traditional science (eg, morality, theology, politics, etc.) when trying to get a broader systems solution, though he did not provide an explicit method for doing so. Certainly, over the ages religious thought and scientific thought have often clashed (and still do today). A Singerian design inquiry sweeps in both in order to learn and make sense of the social system: Churchman (1971). Likewise, many sub-field investigators in the social sciences express animosity toward the views in other sub-fields, yet they each have merit and explain part of the picture. The Singerian inquiry approach to HBM and computational social science can best be supported by computer simulation investigations that attempt to implement many theories into an open process (white box) where scientists, subject matter experts, and analysts can inspect designs, mechanisms, cause and effect, and operation-outcome explanations.

In this view, computational social science is a systems or design topic. It’s not just a black box statistical approach, but instead, a computer-based investigation of the synthesis of theories, expert knowledge, social genomic pictures and trends, and all other evidence and models. Of vital importance to this approach is (1) adopting best-of-breed theories from the social sciences (those that are descriptively valid); (2) keeping an openness to the wide array of systems methodologies and tools, whatever works best for implementing each theory; and (3) a design inquiry approach aimed at learning about a given social system. The point of such a synthesis is to better understand what unexpected effects emerge as a result of policy interventions in network-centric worlds where the social system is complex and poorly understood. This cannot be reliably done in the absence of social science, and not solely with social genomes and black box models. The point is that systems design is the methodological glue that can and must shift the fundamental science in this field.

This is a claim that how we approach the fundamental, underlying social sciences requires a paradigm shift, similar to the metamorphosis happening over the past decade to systems biology. However, it is not a claim that systems models that work in biological (or even mechanical) systems will work in social systems. In general, re-applying such models is a badly conceived idea and they won’t work for long since social systems are fundamentally different from mechanical or even biological systems. They deserve their own form of causality and mechanism modeling. Many authors indicate that agent modeling is promising and potentially revolutionary for modeling social systems – it permits computational experiments and the study of emergence from micro-decision makers.

StateSim is thus a Singerian inquiry system that ‘sweeps in’ models from a range of fields. Yes, it makes its predictions of the effects of actions, but at the same time it is an experiment in how well diverse best of breed theories blend together and serve their synthetic purpose. Section 2 tests this reductive science vs synthesis hypothesis and presents data collected across many StateSim projects.

2) Evolution of the StateSim Tool Suite: Components, Validity, and Applications

As mentioned in the Introduction, many scientists in the human behavior modeling field believe that models should implement accepted wisdom in narrow disciplines (e.g., game theory dictums, solely macro economic theory, or just institutional political theory) and not seek to break with the paradigm and unify across boundaries. A colleague once said: “StateSim is a model of things that don’t belong together.” Hopefully, the reader will recognize this as reductive thinking, and that in the HBM sub-fields, science has lead to many components that work in isolation from each other. My colleague was incorrect. These things do belong together, but reductive science did not tell him (or anyone else) how to do that. That is, when working with best-of-breed theories from different disciplines, it rapidly becomes obvious that they invariably have integration frontiers that have not been addressed as yet. Here we will demonstrate that the parts belong together and in the next 2 sections show how we create a meta-architecture within which to synthesize them.

It is interesting to use science's own methodology to test whether conventional, narrow, and paradigm-preserving, scientific thinking is the best way to model human behavior. Let us use the data presented in this section to see if we can disconfirm the hypothesis: H0: The Singerian, expansionistic paradigm of sweeping in best of breed theories across many disciplines will not provide a realistic (valid) recreation of individual countries. Rejecting this H0 would lead us to H1: The cross-paradigmatic, synthetic approach leads to valid models.

This section serves the dual purpose of overviewing how the StateSim Suite has evolved over 4 time periods as listed down the left side of Table 1, and of using these results as data to test H0. This is far more than a sample size of 4, since in each row there are multiple country models constructed, a variety of scenarios being evaluated, and sometimes 100s of forecasts being validated. The discussion also will highlight selected progress, as the column titles indicate in terms of components created, decision models used by agents, handling of terrain, applications produced, countries modeled and validity and acceptability of test results.

Table 1 – Overview of the Evolution of the StateSim Suite Over Time

| Years | Components Created | Decision Models | Terrain Improvements | Applications Produced | Countries Modeled | Acceptability Tests |
|---------|--|---|---------------------------|---|---|--|
| 2000-05 | Leaders: Value system, reasoning, utility theory | <ul style="list-style-type: none"> •Cognitive Appraisal, GSP •Model of Other | Territory mats | Athena's Prism: World Diplomacy Game | <ul style="list-style-type: none"> • Mid-East • N. Africa • N. Korea | Tested, Accepted, & Fielded at a US Gov Agency |
| 2005-10 | <ul style="list-style-type: none"> •Sociol. of Followers & Factions •Institutional Economy •Dual Sector Theory (White/Black Markets) | <ul style="list-style-type: none"> •Membership/Loyalty •Consensus •Automaton COA | Affordance Theory markups | <ul style="list-style-type: none"> •OneSAF & JSaf •NonKin Village* •DSF Game | <ul style="list-style-type: none"> • Iraq • Afghanistan • Sudan • SE Asia • N. Korea | <ul style="list-style-type: none"> • 3 DARPA Programs • USMC • US AID |
| 2010-15 | Reusable Archetypes Repository | Cellular automata plugin | GIS Shape Files | Metis: Attack the Net | <ul style="list-style-type: none"> • Afghanistan • Levant | <ul style="list-style-type: none"> • 7 Army Schoolhouses • Pentagon/CTTSO |
| 2015-20 | <ul style="list-style-type: none"> •Macro Economics (sectors, debt, S, trade) •The Generator | Gray Zone Strategic Campaign Actor | Region Tool | <ul style="list-style-type: none"> • MEADE • MicroWorld | <ul style="list-style-type: none"> • Baltics • Levant | <ul style="list-style-type: none"> • Pentagon/CTTSO • UK Min. of Defense |

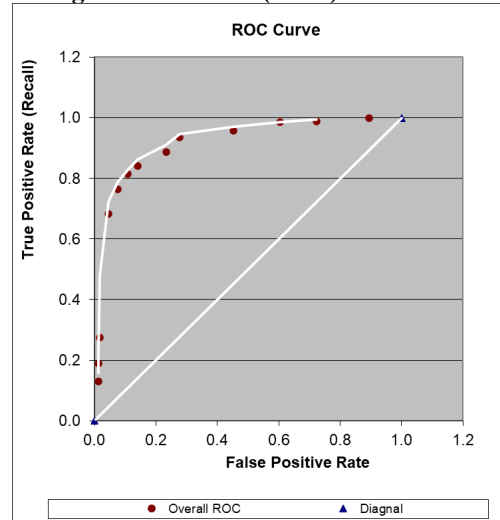
StateSim began in the first part of the 2000s as a model of leader vs leader agents, developed for the Leader Profiling Desk of a US intel agency. The leaders were moral value based, emotive agents, while the groups, followers, and institutions and economy were all handled as virtual 'bingo chips' accumulating on computerized territory mats. This was successfully tested on scenarios for countries in the middle east, Asia, and Africa and it was then fielded in 2005 as a serious diplomacy game called Athena's Prism: Silverman et al (2005). Sections 3.1 and 3.2 below describe how the leaders worked then, and largely still do today. This provides an in depth illustration of how synthetic, expansionistic modeling crosses paradigmatic boundaries and integrates psychology, economics, sociology, personality profiling, and more. One of the tests of the validity of the leader models examined the correspondence of the attack, neutral, and befriending DIME actions that the Thai Prime Minister took toward a rebel group during a 2004 insurgency in the south of Thailand. Using the mutual entropy statistic, $M < 1$, as the criterion to reject the hypothesis that the simulated vs. real prime minister's actions differ, we got a value of $M = 0.045$ and thereby were left with the hypothesis that they were identical: Silverman et al (2008).

As shown in row 2 of Table 1, during the latter half of the 2000s, 3 successive DARPA programs and the AFOSR sponsored us to turn all the factional followers, institutional ministers into autonomous, intelligent agents. We hired a social psychologist to help us add best of breed theories from those fields to the modeling of follower membership and loyalty decisionmaking. We also added two Nobel prize winning economic theories – the Solow, 1954, developing economy or growth model and the LRF Dual Sector theory (Gollins, 2014) that accounts for white vs black markets (and gray). These extensions are described in detail in Silverman et al (2007, 2008). A high level overview is provided in Section 3.1.

In the DARPA effort, StateSim was part of the earlier-mentioned ICEWS team and it went through the same validity assessment effort as the statistical modelers did. We hired country experts and built StateSim models for 4 Asian countries (Thailand, Bangladesh, Sri Lanka, and Vietnam) and trained the models on data from 2002 and 2003. A DARPA contractor then took our models and ran them to predict Events of Interest (coups, rebellions, political repression, etc.) for every quarter of 2004 to 2006. This comes out to 4 countries x 5 EoIs x 12 quarters = 240 backcasts, each of which was an average of 12 runs. Across all these forecasts, StateSim scored accuracy 88%, precision 68%, and recall 84%. The ROC curve of this data is in Fig.1. A **ROC curve** is a **graph** with: The x-axis showing 1 – specificity (= false positive fraction = $FP/(FP+TN)$) and the y-axis showing sensitivity (= true positive fraction = $TP/(TP+FN)$). In general, if the ROC curve is along the diagonal, that is equal to coin flipping. The best possible is a

ROC curve that follows the y-axis and then goes across the top of the graph. The closer the ROC curve approximates that, the better. The ROC curve for StateSim is quite good, though not as good as it was for the statistical model teams. Hence the recommendation to use them for forecasting instabilities and then to only build a StateSim model where one needs to better understand the instabilities and to be able to explore courses of action. It makes perfect sense to use statistical models for early warning and then turn to causal agent modeling for COA analysis.

Figure 1 – Receiver Operating Characteristic (ROC) Curve for StateSim Across 4 Countries



The 3rd row of Table 1 shows two significant events in this time period. The first is the creation of a reusable artifact repository. We postpone discussing the significance of this event until Section 4 which concerns speeding up country model creation. The second event is the Metis: Attack the Net application built for the US Army PEO STRI. This embedded StateSim behind a geographic display oriented military command and control system. StateSim was used to drive the activities and performance of the socio-economic and governmental institutions and to drive behaviors of a population of 100,000 in 3 towns and 2 rural regions all in Afghanistan. In Metis, StateSim also drives the behaviors of a network of about 150 insurgents. If one examines the 3rd column of Table 1, one can see the improvements made over time to handling of terrain. In early versions of StateSim, the agents could only reason about the physical world if it was hard-wired into large Markov chains about states the world objects could be in. These were ever-growing and difficult to maintain, so we switched to psychological affordance theory. Rather than agents having to maintain knowledge about all the artifacts in the world and their possible states, this meant the artifacts themselves would carry all the knowledge any agent needs to know about them. This made it far easier to maintain and allowed our agents to move between worlds, plugging into many 3rd party software systems listed in rows 2 and 3 of Table 1: eg, see Silverman et al (2019) for full detail. With the Metis application, battalion command staff could train on attack the net planning for 10,000 blue forces that they would send out to detect, disrupt, and defeat the insurgents. They never realized they were utilizing StateSim which operated all the blue, red, and green forces, as well as the civilian population. This system was fielded at 7 Army school houses following 7 different 3-day long acceptance tests - one at each site. These acceptability tests are something akin to a Turing test where the users decide if the agents in the model are as realistic as their realworld counterparts. This is thus yet another example of rejecting H0 and accepting H1 – the synthetic or trans-disciplinary approach. We don't know how we could have passed these acceptance tests otherwise.

As the 4th and final row of Table 1 shows, we worked over the past half-decade to update StateSim's economic models to handle developed and not just developing world economies. Our current sponsors are less worried about insurgencies in weak nations, and more focused on revisionistic adversaries seeking to disrupt alliances and world order amongst the developing and developed nations. We worked to extend StateSim's dual sector economic theory modeling by adding in models of other sectors (several private ones, several public ones) as well as the money supply, debt and banking, and balance of trade.

As the 2nd column of Table 1 shows, up until now, StateSim has gotten its excellent validity scores and passed many realism acceptability (or Turing) tests largely using leader and follower agents that are 1 step ahead, reactors. The agents of this ilk, as shown in Sections 3.1 and 3.2, are sensitive to their own and others' values, relationships, quality of life and so on. They are not at all simplistic. However, they are just muddlers thinking but a step ahead. This

is an interesting confirmation of the ubiquity of Lindblom (1959)'s "Muddling Through" or incrementalism theory. To create revisionistic actors, we had to expand leader reasoning so the agent considers campaign steps, progress in reaching the next step, and shifts in action choices as the campaign proceeds. We have succeeded in creating two such leaders by modifying the reactive leaders to allow them to consider rewards across time steps. StateSim now has an economic campaign agent as well as a territory campaign agent, though both are still evolving. That is, one type of revisionistic adversary is largely an economic force, adept at creating economic dependencies and debt traps. Their purpose is less to gain territory as it is to gain economic advantage by building an economic empire. Another type of revisionistic actor is one more focused on gaining and controlling territory by using instruments such as trade imbalance, force posturing, misinformation campaigns, and the help of proxies inside a country to carry out political and economic disruption, thereby weakening the current government until regime change can be effected. Testing of these is ongoing, however, two tests are summarized in Sections 4.2 and 4.3. In both cases the users have indicated the campaign agents are passing the Turing or acceptability tests thus far.

In this section, we proved the value of the synthetic paradigm without fully clarifying what it entails. The next two sections (Sect 3.1 and 3.2) are included to show how the synthetic paradigm works. Specifically, the StateSim approach to synthesis is to adopt a higher level decision-theoretic mathematics that comprises a sociological game theory notation and to map social science theories (best of breed components) into it. These sections were written in 2007 (although references are updated) and describe exactly how the leaders work in rows 2-4 of Table 1: see Silverman et al, 2007. The leaders in row 1 are similar, but somewhat simpler. The basics of StateSim have not changed since then. Section 3.1 gives a high level overview of the StateSim mathematics, while Section 3.2 explains leader decision making. The primary change affecting Sect 3.1 is that we have extended the economy model as indicated in Table 1, row 4. But that is not relevant to this discussion. Also, Section 3.2 describes the cognitive model using 1 step ahead utility maximizers (Muddlers). In addition to this, as the 2nd column of Table 2 shows, we also have several other types of decision algorithms such as:

- Consensus model – For modeling councils of tribal elders, each is a leader model as explained in 3.2 and their decisions are combined using an influence weighting function to reflect their individual status.
- Follower agent model (Membership/Loyalty) – In the same vein as Sect 3.2, several best of breed models of loyalty, membership, grievance/sentiment, and relations are synthesized: see Silverman et al., 2007. Followers make group membership and action decisions (eg, join, exit, protest, fight against, etc).
- Automaton agent – These are finite state automata that follow a "course of action" (COA) script. Such agents run the macro-economy model, the ministers of the institutions, various external actors (eg, NGOs, bordering countries, climate, etc), and are the user's avatar to insert COAs and study their impact.
- Cellular Automata Plugin – Cellular automata, or grids of simple finite state automata, are useful to model the propagation of information and sentiment across space if a given scenario needs this.
- Gray Zone (GZ) Strategic Campaign Actor – This converts the same leader agents (sect 3.2) from 1-step ahead utility maximization to a multi-step campaign strategist. Individual step utility is replaced with multi-step or campaign utility and thresholds of outcome success of individual steps are used to monitor progress toward the campaign goal. This agent is described further in Sect. 4.3.

These agent algorithms are not described here, but the interested reader may find them in the various other papers on StateSim over the years. The most recent extensions to the Economy model and the addition of the GZ actors is described in the User Guide in the ancillary materials. Also the GZ actors are further discussed in Section 4.

2.1) Profiling and Sociological Game Theory

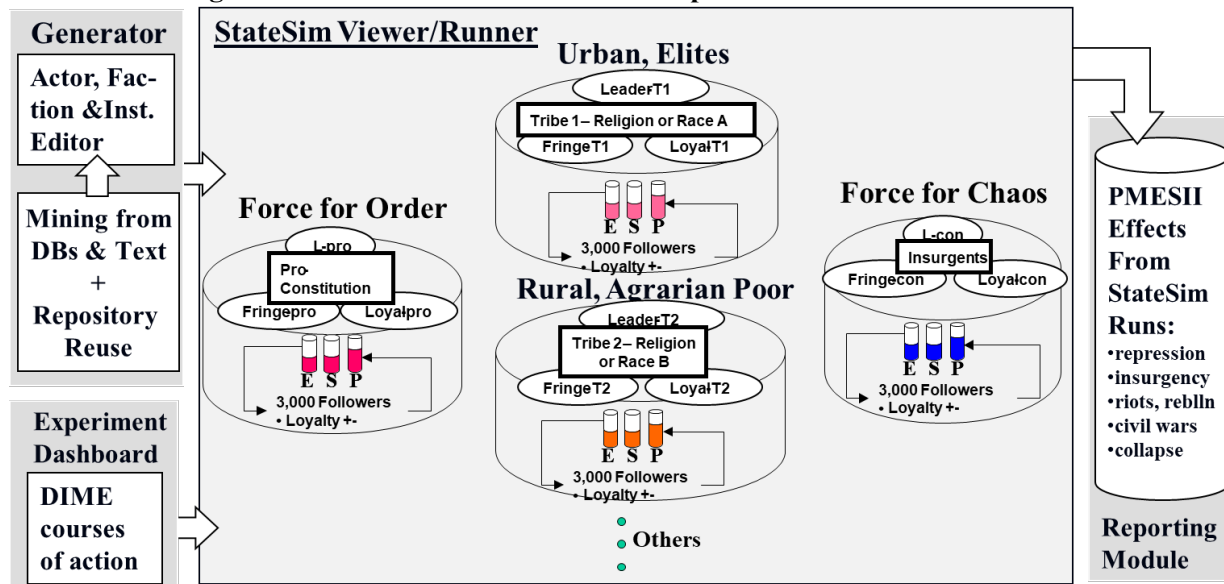
StateSim facilitates the codification of alternative theories of leader, faction, and institutional interaction and the evaluation of policy alternatives. StateSim is a tool where you set up a conflict scenario in which the factional leader and follower agents all run autonomously. Generally, you are the sole human interacting and using a set of DIME actions to try and influence outcomes and Political, Military, Economic, Social, Information, and Infrastructure (PMESII) effects (see Figure 2).

Factions are modeled as in the center of Figure 1 where each has a leader agent, two or more archetypical sub-faction follower agents (core and fringe), a set of starting resources (Economy, E, Security, S, and Politics, P), and a representative set of over 1,000 follower agents easily scaled to the size of the actual group. For fighter groups (rebels, insurgents, etc), the fringe might be opportunists while the core are true believers. For civilian groups in the most recent version (row 4, Table 1), StateSim now scrapes the cloud and instantiates several follower archetypes, each representing a demographic sub-faction (see Sect. 3).

A leader is assumed to manage his faction's E- and S- tanks so as to appeal to his followers and to each of the other tribes or factions he wants in his alliance. Each of the leaders of those factions, however, will similarly manage their own E and S assets in trying to keep their sub-factions and memberships happy. Followers determine the level of the P-tank by voting their membership level (see Silverman et al, 2007). A high P-tank means that there are more members to recruit for security missions and/or to train and deploy in economic ventures. So leaders often find it difficult to move to alignments and positions that are very far from the motivations of their memberships.

Despite efforts at simplicity, stochastic simulation models for domains such as this rapidly become complex. The strategy space for each leader facing only two other leaders is in the trillions of options, a number impossibly large to explore. As a result, StateSim's Experiment Dashboard (left side of Fig.1) permits inputs ranging from one course of action to a set of parameter experiments the player is curious about. On the bottom left is the profile editor of the personalities for the leaders and sub-leaders, and of the key parameters that define the starting conditions of each of the factions and sub-factions. There is also a repository of reusable archetypes – ie, reusable agent leaders, followers, and institutions. This is used to speed model construction as will be explained in Sect. 3.

Figure 2 – Overview of Models and Components of the StateSim Suite



Game Analysis

StateSim runs a set of multiple games, $G = \{G1, G2, \dots, Gn\}$ simultaneously. Within a faction one may observe games between rival leaders, between leaders and followers, and follower on follower. The across-faction games include attempts to cooperate and/or compete with other factions' leaders and followers, and/or attempts to contain factions aimed at your own downfall. For discussion's sake, consider these as iterated semi-cooperative games (ISCGs). This game formulation is the simplest game one can analyze involving conflicts between (and within) factions. Using it helps to clarify many of the key elements of these conflicts.

Let us next consider how StateSim's games might be treated by two types of ISCG agents, namely:

Rational Actors: Presumed normative and devoid of psychic concepts as in post-WW II economic theory and intro game theory classes - perfectly informed, purely logical, constant discount rate (i), and motivated by self-interest to maximize their material payoffs. All actors have identical payoff functions where they compute $R\{E|S|P\}$ as uncontested resources, Q as resources at stake, and $CstA$ as the cost of actions. The expected payoff is thus: $Payoff = R + Q - CstA$. Mutual conflict or fight-fight is a well-known Nash equilibrium. We know also that if $CxCy > FxCy$, then mutual cooperation is Pareto optimal and in repeated games, if the agent histories are remembered, no agent is excessively powerful, and agents start with mutual cooperation, then the following is the well-known mixed strategy that will prevail: attack if provoked (tit-for-tat) to deter other leaders from taking advantage, but otherwise cooperate. The subgame perfect equilibrium consists of long periods of cooperation punctuated by occasional conflicts. Ignoring rare conflicts, one may write the predicted payoffs for any given 'rational' agent in alliance with others as:

$$\text{PAYOFF}_x = \sum_{t=0}^T C_x C_y(t) / (1+i)^t \quad (1)$$

Descriptive agents: Swedberg (2001, p.325) states "If sociological game theory is not to end up as an artificial exercise, it is absolutely essential that the beliefs, ideas and experiences of the actors themselves are moved onto center stage". One must profile the individuals involved to find out the inventory of items at stake and to build realistic agent models. We do this with best-of-breed social science instruments (e.g., see Sect 2.2). Such actors use these approaches to decide everything from R and Q, to the size of an action, to how to discount (i), to how much they are willing to pay for their gambits (CstA), etc. – one wouldn't even expect to use the same formulas for normative vs. descriptive computations. Aside from material payoffs, these agents attend to moralistic issues driven by their emotional value (emV) and how relationships ($\square K_{xy}$) change; they may commit errors and use biased heuristics; and they may see games through a different lens (e.g., settling grievances, fast track to next life). For these agents, the payoff function becomes:

$$\text{Payoff} = R + Q - \text{CstA} - |\square K| + \text{emV}. \quad (2)$$

Researchers like Askari et al., 2019, show that as one alters agent aspirations (something equivalent to eliminating emV and $\square K$), the stable equilibrium [1] collapses and the prediction of fight-fight becomes near-continual. Burns et al. (2017) show that divisibility of Q as well as emV and $\square K$ type issues must be elaborated if scenarios are to be properly profiled and for conflicts to be settled. Results like this mean that one must reduce the guess work about what drives the resource disputes, moral dilemmas, and social relationship grievances.

2.2) Leaders: Synthesizing Cognitive Appraisal and Personality Profiling

Hypothesis H1 of Section 2.0 states that social science models are improved by synthesizing best-of-breed component theories. There are many StateSim papers showing how to do this. For the sake of completeness, here we illustrate how to do it for the emV term of Equation (2). Social science theories rarely come with equations and invariably are operationally underspecified. This section thus explains a way to convert qualitative theories into quantitative methods and use them to fill in what is missing from other component theories. Specifically, in order to compute emV, we will have to operationalize, implement, and synthesize decision-theoretic versions of cognitive appraisal theory, Bayesian-weighted value trees, and several instruments for personality and culture profiling. The resulting synthesis is what StateSim has been using for 20 years and across all the rows of Table 1.

Profiling of personalities has not yet reached the stage of a mature science with first principles; however, there are best-of-breed profiling instruments with respectable field trials and high inter-rater reliability. These are useful for creating agent frameworks with greater degrees of realism. Such implementations, if done carefully, may in fact improve profiling science. Hendrickson and McKelvey (2002) suggest that social science theories, in general, need to be computationally formalized as agent models to show they are analytically adequate. These models in turn must be subjected to correspondence tests against real world phenomena to verify them (ontologic adequacy). This two step testing process improves the science by revealing the agenda for advancement.

Unlike the evolutionary tradition where personas are mutated, this approach of profiling real personalities within connectionist agent models allows one to watch the generative mechanism and to observe what they do, how they learn and adapt, and what macro-behavior emerges from the actors' micro-decisions. Using profiling instruments reduces the dimensionality to the traits and factors they require, and where these are applied, we can use training datasets, fill in the traits and factors of archetypical as well as real characters, conduct validation tests, and treat these parameters as no longer independent variables clouding the larger political analyses – they exist within encapsulated components and only their inter-relationships to other parts are significant when assessing the whole. This is no different than systems engineering for any complicated device. A crash test of an automobile does not depend on how the pistons fire. We similarly encapsulate other parts of the faction model – e.g., the (E|S|P) resource tanks that we currently model as stacks of poker chips that grow or fall. One can plug in finer resolution models for any given tank without affecting overall system performance. With that in mind, we turn now to the best-of-breed profiling theories we implemented as leader and follower models.

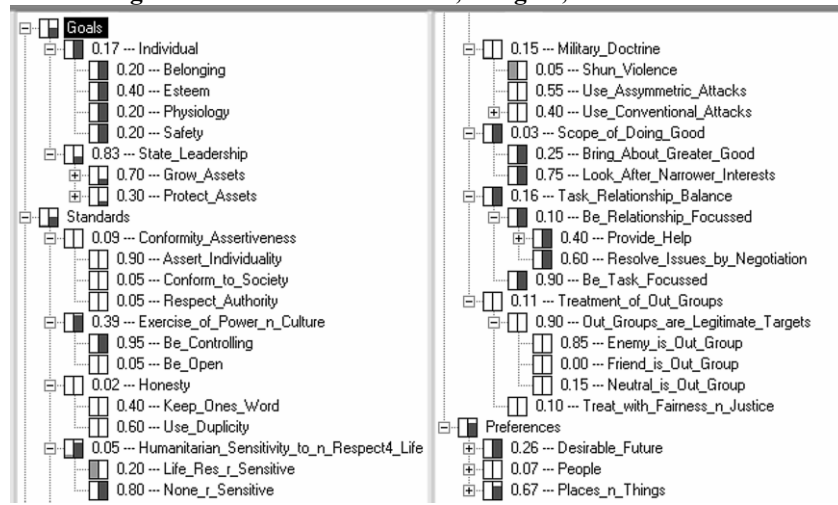
Cognitive appraisal theory is the way psychologists say that people assess a situation against their values to determine how they feel about it and to decide what to do about it. In StateSim, each leader and follower is modeled within a framework postulated by Ortony, Clore, and Collins (1988) – referred to as the OCC framework. Despite the reference having no equations and only 1 flowchart, this OCC framework is the most widely implemented cognitive appraisal model in the agent world and StateSim has a full implementation. The framework indicates that actors make decisions by appraising the world against their value trees and choosing the best action as the one that maximizes their

values. OCC also specifies three value trees called Goals, Standards and Preferences (GSP) trees. These are defined, but unspecified and no examples are provided.

StateSim implements these as multi-attribute value structures where each tree node is weighted with Bayesian probabilities or importance weights. A Preference Tree is one's long term desires for world situations and relations (e.g., no weapons of mass destruction, stop global warming, etc.) that may or may not be achieved in the scope of a scenario. In StateSim agents this translates into a weighted hierarchy of territories and constituencies (e.g., no tokens of leader X in resource Y of territory Z). The Standards Tree defines the methods an individual is willing to take to attain his/her preferences, and what code that others should live by as well. Finally, the Goal Tree holds short term needs the agent seeks to satisfy each turn (e.g., vulnerability avoidance, power, rest, etc.). StateSim's GSP tree is a value model editor that allows one to (a) implement leader and follower profile instruments as nodes on the trees and (b) set the weights on the nodes which in turn implements a personality profile (see Figure 3).

The question is what to use as the nodes (individual values) on the GSP trees. In Standards one could map the bible, military doctrine, and civilian laws, but this is unending. Likewise Goals and Preferences can be equally long. The problem is to balance between a minimal set of nodes and enough descriptive power to be able to profile the values of specific leaders and archetypical demographic follower agents. We restrict the remainder of this discussion solely to how to do this for leader models.

Figure 3 – GSP Tree Structure, Weights, and Activations



Perhaps the best leader profiling instrument is Hermann (1999) who offers a descriptive theory of leader style that is measurable and can be fully implemented in this framework. After two decades of studying over 122 national leaders including presidents, prime minister, kings, and dictators, Hermann uncovered a set of leadership styles that appear to influence how leaders interact with constituents, advisers, or other leaders. Hermann determined that seven traits are particularly useful in assessing leadership style: need for power, control, openness (combines 2 traits), task vs. relationship focus, distrust, and in-group bias.

In Hermann's profiling method, each trait is assessed through content analysis of leaders' interview responses as well as or other secondary sources of information. Hermann's research also has developed methods to assess leadership at a distance, based mostly on the public statements of leaders. Hermann (1999) has developed mean scores on each of the seven traits. A leader is considered to have high score on a trait, if he or she is one standard deviation above the average score for all leaders on that trait.

In order to implement the Hermann instrument as an agent model (GSP Trees), we had to do the following: Need to increase power (and its inverse, protection) is both a long term Preference as well as a short term Goal. So it appears on both trees. In the Machiavellian and Hermann-profiled world of leaders, the goal tree reduces to a duality of growing vs. protecting the resources in one's constituency. Expressing goals in terms of power and vulnerability provide a high-fidelity means of evaluating the short-term consequences of actions. Most of the other Hermann traits govern personal and cultural norms and thus appear on the Standards tree.

The UN GLOBE model of leaders (House, 2004) includes several traits like Hermann's but also adds Scope of Doing and Sensitivity to Life (humanitarianism) which we adopt here as well since these are central to how our leaders perform in the StateSim game of games. These games also concern how agents grow vs protect assets and

what their values are concerning military actions. So those are the other GSP nodes added to the profile to capture the doctrine a leader adheres to when considering his Economic and Security tanks. Beneath each subnode that has a + sign, there are further subnodes, but under the G- and P-trees these are just each faction's resources in each territory with valence and importance valuated weights.

The structure of the GSP trees is shared by all StateSim agents. However, the tree weights are unique for each agent and thus capture individual differences that may be determined by culture, ideology, or personality. Other papers discuss how the weights may be derived so as to increase credibility: e.g., see Bharathy & Silverman (2013), Silverman et al. (2007). An example of the weights is the insurgent leader shown in Figure 3. He is low on conformity, humanitarianism, scope of doing good, and treating outgroups with fairness, and high on exercise of power, and asymmetric warfare.

GSP trees are used by the agent for all decisions – e.g., selecting a next game action, determining faction alliance moves, or deciding on a speech act. They give each agent a robust and individual worldview. When contemplating a decision, the agent calculates the subjective expected utility (SEU) it expects to derive from every action available to it, as constrained by perception and chooses the alternative that maximizes SEU. Thus

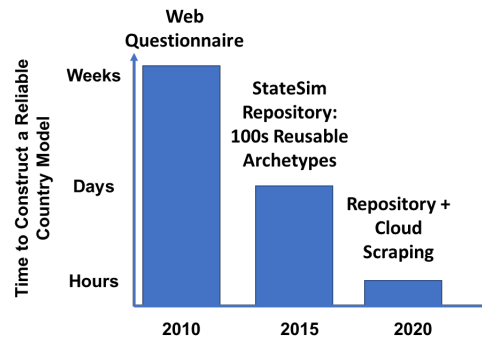
$$\text{Best Response (SEU)} = \text{Max}\{\sum U(a_k) * \text{Pr}(a_k) * \sum (r_j) + \dots\} \quad (3)$$

where utilities (U) for next actions, a_k , are the anticipated E|S|P tank gains or losses the actions afford combined with how those affect the nodes of a given agent's GSP trees. $\sum (r_j)$ is a function that captures the strength of positive and negative relationships one has with agent or object j that are effected or spared by a_k , and \sum handles merging and discounting (decay) prior GSP activations. Probabilities assess the likelihood of success or failure. Also, the GSP tree weights adhere to principles of probability; e.g., all child node insights add to unity beneath a given parent, activations and weights are multiplied up a branch, and no child has multiple parents (independence). Although we use fixed weights on the GSP trees, the succeed and fail reservoirs on each node (see Fig.3) serve to render them dynamic and adaptive to the agent's current needs. Thus, when a given success reservoir is filled, until \sum decays it, that tends to nullify the importance of the weight on that node (or amplify it if the failure reservoir is filled). In this fashion, one can think of a form of spreading activation (and deactivation) across the GSP structure as a game proceeds.

3) Speeding Up Model Creation

A major obstacle to analysts for using StateSim has been the lead time to set up a model of a country or region of interest. Often when crises arise, the analysts have only a few hours before their bosses want a plan showing alternative courses of action and the likely outcomes of each alternative. As Figure 4 shows, we have only recently achieved that level of performance. The horizontal axis shows the era, while the vertical axis shows the typical length of construction time. In this section we review why model building was so time consuming initially, and what we did to improve the situation. The next section provides data showing how fast the generator now works.

Figure 4 – Improvements in StateSim Model Construction Time



As the lefthand bar on Fig 6 shows, in StateSim's early days we succeeded in producing an 80 table web questionnaire that collected all the StateSim model parameters from a Subject Matter Expert (SME). Our approach was to pay 3 SMEs on any given country to complete the questionnaire. This is something a true SME (eg, country desk expert) could do in 3 to 5 days, but which would take others 2 or more weeks to do. Also, we would triangulate the answers of the 3 SMEs and cross check against about 4 dozen databases on the web. We would then build the country model, run it, and test it against training data (past scenarios) in the country of interest. Overall this process

took weeks for any given country model. We got excellent validity scores during the DARPA ICEWS horse race (see earlier Fig.1), but only completed models of 4 countries. During this same interval, the statistical modeling group, using only a handful of variables, was able to model and predict outcomes for all 200 countries of the world.

We do not expect to ever beat this. Statistical prediction models will always be faster and more accurate than causal models. But the former is a black box and one needs the human behavior models if one wants to explore alternative COAs, counter-factuals and what-ifs. Analysts should use the statistical models to forecast where crises are likely to occur around the world and then construct a causal StateSim model of the country in crisis to better understand causes and to explore intervention and stabilization ideas. To make this tandem approach work, however, we knew we had to reduce the weeks of construction time by one or two orders of magnitude.

Serendipitously, after a few years of constructing 18 or 20 country models the slow way, we had accumulated a large number of validated agent models (e.g., generals, dictators, insurgents, politicians, etc), group models (ethnic, religious, civil, etc.), and numerous institutional models (eg, health, education, security, criminal, etc). In fact we had numerous versions of each of these (e.g., 6 insurgent leaders, 8 military dictators, etc.). Most of these were named individuals, groups, and/or institutions that had been created by paying SMEs to profile them and by testing and tuning them against actual training data. After a while, it became apparent that there are only so many variants in the world, and a benevolent leader or harsh and corrupt autocrat model could be reused. Our Al-Baghdadi model was constructable from our al-Zarqawi model. All that was needed was a taxonomy (roles, groups, and institutions), a repository of the reusable artifacts keyed to the taxonomy, and a case based reasoning engine to link the current country model needs to the already constructed artifacts from the repository. We constructed this and it is described in Silverman et al. (2016). This approach did help reduce construction time by an order of magnitude – from weeks to days. However, alone it was not enough.

In order to further speed up model construction, we are also now scraping data directly from numerous leading political and social science databases so that many of the key parameters of a country model are “pre-populated” rather than built from scratch by expert judgments each time. Table 4 provides an overview of some of the principal data sources we extract from. In the economic sphere, we extract the size of six sectors of the economy – black market (from Havoscope) and gray market (IMF); three private sectors of each country from World Factbook (agriculture, industry, and services); and lastly the public sector from several sources including five public institutions (health, education, public works, security, and law). The latter helps to prepopulate quality of life information sorted by faction and demographic groups like labor rates, unemployment, tax payments and savings, as well as public goods and services allocated.

In the political and ethnic spheres, we have integrated sources such as the Polity IV database (Center for Systemic Peace, 2020) to measure regime type, the UCDP/PRIO Armed Conflict Dataset (Pettersson and Oberg, 2020) to capture the state of conflict in a country, the Ethnic Power Relations (EPR) database to break down the key ethnic or religious groups in a country and the balance of power between them (Vogt et al, 2015), and the World Values Survey (WVS) to obtain information on public preferences and demographics in each state (Inglehart, et al 2014).

Two of these efforts merit particular attention, one at the group level and one at the individual or micro-level. First, the breakdown of a country into ethno-political groups which dynamically interact with one another has always been one of the core features of StateSim. We now automate this task with data from EPR, which is a leading resource in the study of conflict and instability based out of ETH Zurich and codes the “politically relevant” identity groups in each country around the world as well as their political power and authority within the country. Using this information, StateSim now automatically “loads” the salient ethnic or other identity groups in any user-selected country along with their share of political authority, creating the foundation for a realistic conflict simulation with ease. A similar process is followed for the key political parties (which are extracted from the World Bank’s Database of Political Institutions) and rebel groups (extracted from the Big, Allied, and Dangerous dataset) active within a country as well. With some simplifying assumptions that adjust the authority of more formal/institutionalized vs. informal/militarized actors based on a country’s level of democracy in the Polity IV database, we are then able to present a useful baseline picture of a country’s major ethnic, political, and military actors automatically before any user input is needed.

Second, we have also integrated StateSim meaningfully with the World Values Survey (WVS), a top source of individual-level public opinion data around the world over the last four decades. In particular, we have linked the WVS data by group with all of the actors in the EPR database so that StateSim can break down identity groups (e.g., Sunni Arabs in Iraq) by their demographic and attitudinal characteristics. In practice, this means much richer modeling automatically extracted and ready “out of the box” for the user. In particular, users can select up to three demographic attributes (from age, gender, education, income, urbanity, and ideology) and StateSim will extract the shares of each of these factors for every group in the model, allowing them to act as semi-independent follower factions underneath their broader group umbrella (for example, poor urban male Sunni Arabs in Iraq). These fine-grained breakdowns are also available in the domain of attitudes: we have mapped questions in the WVS questionnaires to all of the values in

the GSP trees, so groups and sub-groups have distinct worldviews that are pre-populated as well. This integration with leading political and social science databases thus helps tether StateSim more to high-quality social science data and enrich its modeling, while helping achieve the primary goal of speeding country model generation.

Table 2 – Currently Used Open Data Sources: Growing List

| Overview | Political | Conflict | Economic |
|--|--|--|---|
| <ul style="list-style-type: none"> • The World Factbook • Quality of Government • World Value Survey • OECD Database | <ul style="list-style-type: none"> • The Polity Project • Ethnic Power Relations Dataset • World Bank Database of Political Institutions • State Fragility Index and Matrix • List of Ruling Political Parties by Country • Center for Systemic Peace • World Justice Peace Rule of Law Index • World Bank Government Effectiveness Index • Sustainable Society Index | <ul style="list-style-type: none"> • UCDP/PRIO Armed Conflict Dataset • Big, Allied, and Dangerous • List of Ongoing Armed Conflicts • World Population Review | <ul style="list-style-type: none"> • International Monetary Fund Economy Classification • International Labour Organization • Havoscope Global Black Market Information • The World Factbook • OECD Global Revenue Statistics Database |

All of the relevant material for each country is scraped from the full set of sources ahead of time and stored in a database, a snapshot of which is available to the StateSim Generator. At Generator launch, when a user selects a country to model, the data for the relevant model parameters are instantly available and instantiated into the model's parameters. The utility of this parameter initialization lies with its ability to speed up the model creation process and improve the accuracy of the model. We have found this data-driven approach to be acutely helpful when it comes to dealing with salient parameters whose essence is difficult for users to quantify such as the GSP tree differences between demographic sub-groups, or the myriad of economic flows across institutions and groups, among others.

4.1) Tests of Construction Speed and Model Acceptability

Over the past 2 years, we have conducted several tests of the Generator. The tests involved sessions with different information operations analyst groups in the UK Ministry of Defense (MoD). In each session subjects were oriented/trained in the StateSim Suite, used the Generator to model a country of their choosing, and then ran the base case (as-is) as well as at least 1 what-if Course of Action (COA) scenario to see how they could alter the outcome. Subjects often worked alone or in pairs. The test outcomes are summarized in Table 3. The 3 tests in 2019 took t to 1.5 days each (8-12 work hours). The groups accepted StateSim as reasonably realistic and for follow-on funding subject to specified extensions. These extensions were to fix numerous bugs, bottlenecks, and ways to further improve the Generator screens and scrapings. The most important extension the subjects wanted was to be able to have the data available on any ethnic or political faction's demographics. Fortunately, the World Value Survey (WVS) provides this for roughly 120 countries. We completed all the requested extensions and conducted another test with a new group of users. The result reveals another order of magnitude improvement in time required to build and run StateSim country models and lead to an invitation to integrate StateSim into their command center. The final row of Table 3 shows the lead author's recent usage of the Generator and Runner. The actual model inputs and outputs are summarized in the next section where the readers can see for themselves if it is acceptably close to real world scenarios.

Table 3 – Results of Recent Timing and Acceptability Tests for the Rapid Model Generator

| Year | Subjects (n) | Country Modeled | Ran Base Case | Ran COA Scenarios | Time Needed To Build & Run | Pass Acceptability Test? |
|------|----------------------------|-----------------|---------------|-------------------|----------------------------|-----------------------------|
| 2019 | UK Ministry of Defense (4) | Brexit | Yes | Yes | 12 hours | Yes, funded s.t. extension |
| 2019 | UK Ministry of Defense (3) | Estonia | Yes | Yes | 12 hours | Yes, funded s.t. extension |
| 2019 | UK Ministry of Defense (6) | USA | Yes | Yes | 12 hours | Yes, funded s.t. extension* |
| 2020 | UK Ministry of Defense (3) | Estonia | Yes | Yes | 2.5 hours | Yes, integration planned |
| 2020 | 1 st Author (1) | Belarus | Yes | Yes | 2 hours | Yes (see Sect 4.1) |

* - Extension required to add the World Value System (WVS) breakouts of demographic sub-groups.

4) Course of Action (COA) and What-If Case Studies

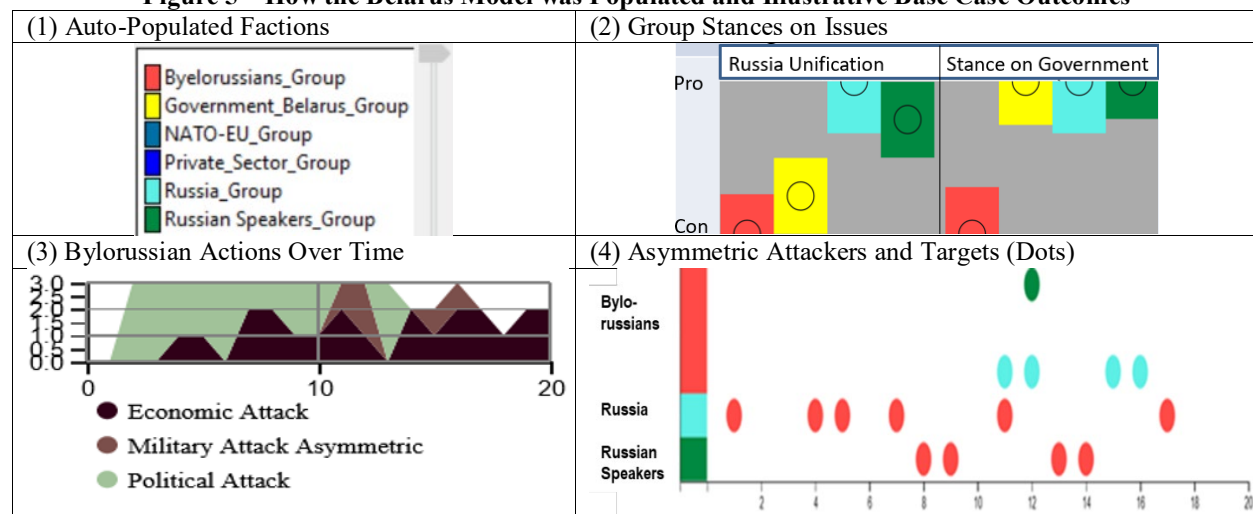
This section includes 2 case studies that illustrate three things: (1) how well the Repository and Generator do in facilitating the creation of scenarios involving reusable StateSim agents, groups, and institutions; (2) the behavior of the new Gray Zone agents; and (3) the use of COAs that alter scenario outcomes. The first case study was created in September 2020 and its the Belarus model that was just shown in the last row of Table 3. It shows what the Generator produced for that acceptability test and it projects what is going on in that country and one way that NATO-EU might influence outcomes. The second case study was created as part of MEADE during the summer of 2018 with the help of the StateSim repository (the Generator was only a prototype then) guided by about 2 days of effort from a Latvian SME. According to a separate Baltics SME hired after the runs, it illustrates realistic behaviors of the Gray Zone actor (Russia) as the US and Germany waiver and then withdraw support for Latvia. The SME stated it gives the same advice as he would provide, only with more details.

4.1) What Are the Primary, Secondary, and Tertiary Effects of COAs?

The point of this experiment was to see if a plausible model of a country could be assembled and run in an afternoon. We allowed 1 hour to read about the country and construct the model. The country chosen was Belarus which is undergoing internal conflict due to its leader refusing to step down despite losing the election in August 2020. The Generator indicated the country was an autocracy and also is in a “cold war” struggle with Russia which wants to draw it away from the EU. The Generator scraped the cloud and then populated the model with the factions shown in Fig 5.1. Byelorussians are about 87% of the population, while Russia and NATO-EU are external agents, with Russia using the autonomous gray zone decision model and NATO-EU is an automaton that requires a COA script to follow.

The Generator filled out many of the other model parameters, but a few required user input. One such was the issues and stances on the issues as shown in Fig 5.2. We entered 2 issues – unification with Russia and stance on the current regime staying in power. The Byelorussians are against both and the current regime is friendlier with Russia but does not want reunification.

Figure 5 – How the Belarus Model was Populated and Illustrative Base Case Outcomes



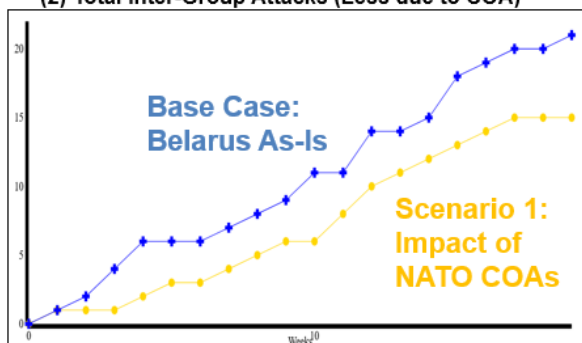
Running the Base Case or As-Is model for 20 weeks showed in Fig 5.3 that the Blorussians are unhappy with the outcome and are instigating political (protests, demonstrations, etc) and economic (strikes, boycotts, etc) attacks. This is completely accurate and is exactly what the news is reporting is happening in the country right now. Also, Fig 5.4 shows the Byelorussians are suffering a series of assymetric (small) attacks from the Russians and Russian Speakers. This is consistent with Russia’s typical modus of inserting “little green men” and use of proxies (the Russian Speakers in this case) to sow discontent and to foment strife. It leads to the Byelorussians making a few counter attacks later in the run. The one thing missing from the run is that the Government (yellow) is known to be arresting demonstrators and detaining and torturing them. Aside from that omission, this is a reasonable picture of what the participants seem to be doing on the streets in Belarus. With just a few minutes tweaking of the government leader agent’s archetype (in the Generator) this could become a robust model of the situation.

NATO-EU is not likely to stand by silently. But what to do? Fig 6.1 shows an illustrative COA of things one might expect NATO-EU to try. This is not an illustration of the tool to build the COA, but is an output of that tool showing what COA was built and run. The COA is called “Aid Protesters” and it has 6 actions over time. The first three actions or rows are to help the Byelorussians with economic aid, moral support and health care jobs (in the pandemic). The amounts (in \$millions) and intervals are shown on the right. In addition to carrots, there are 3 “sticks” including posturing and decreasing aid to the Belarus Government as well as a media propaganda campaign against Russian interference where blame is assigned to the Russian Speakers (even though this is a NATO-EU campaign). The results of these COAs show a drop in inter-group attacks (Fig 6.2) and some minor shifts in group power (Fig 6.3) both in the desired direction. These 3 charts of Fig 6 may be thought of as illustrating the primary, secondary, and tertiary impacts of the actions. Given the agents have emotions, social relations, health, etc, there are more charts one can dive into to try and more fully see the effects. We cannot do that here given space limits. Likewise, there are many other COAs one can experiment with to study how participants might react and how to stabilize the region. But again, we limit ourselves to just one example.

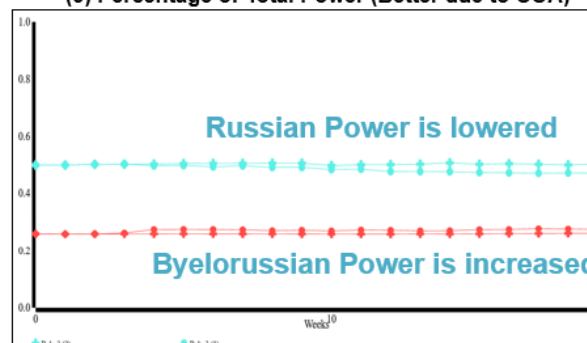
Figure 6 – Cross-Run Comparison: Base Case vs. NATO Intervention
(1) NATO-EU’s Course of Action (COA) Schedule



(2) Total Inter-Group Attacks (Less due to COA)



(3) Percentage of Total Power (Better due to COA)



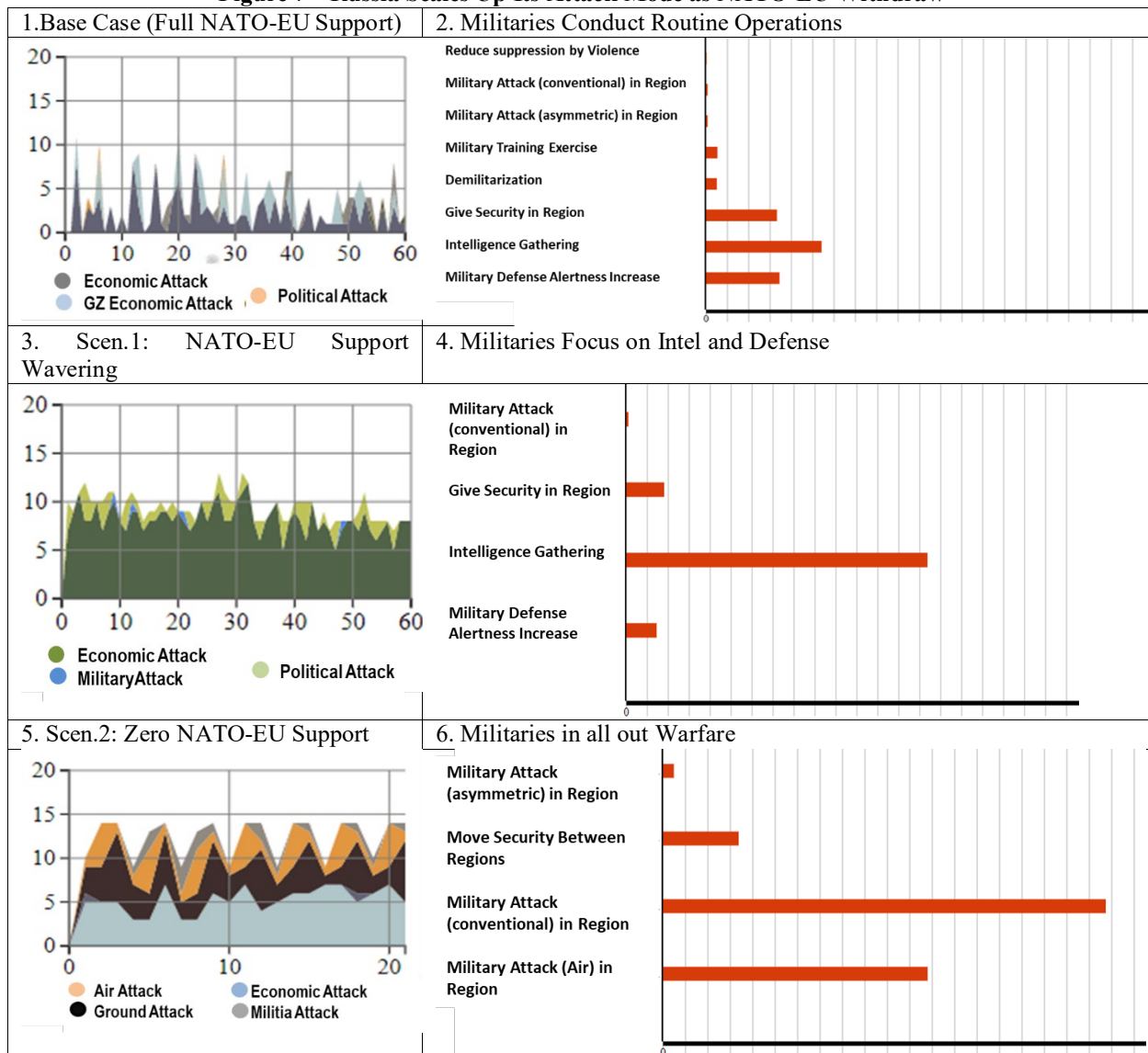
4.2) Dynamics of Gray Zone Actor in Latvia

The purpose of StateSim is to allow analysts to profile agents, drop them into a context, and watch what they do to each other. In this experiment, we modeled the groups in Latvia, another Baltic country that used to be in the USSR. The model is not unlike that of Belarus, except Latvia is a democracy that is affiliated with the EU. There is a minority right wing, populist faction that prefers independent nationhood and a minority faction of Russian-speakers (the Pro Russian Left) who could act as internal proxies for Russia’s reunification interests. The majority of the population are center-right, Latvian natives who enjoy democracy and modern alliances. Russia is an external Gray Zone (GZ) agent with some internal agents assisting inside Latvia. The Latvians tend to treat the Pro Russian Left as a lower class group, so they are a natural to serve as Russia’s proxy as their counterparts did in the Ukraine.

This experiment includes 3 runs: a base case and two alternative scenarios. In the base case run, Latvia has the full support, protections, and aid from NATO-EU and the US (automaton agents providing aid). As seen in Fig.7.1, the vast majority of attacks are economic and political. These are largely either various groups competing and/or the populists going against the Pro Russian Left. The lightest color spikes are Russian attempts to weaken the Latvian economy (eg, oil price hikes or supply interrupts). The various militaries in the scenario (Latvian government, militias, and Russia) are conducting only routine operations (Fig. 7.2) such as patrolling, defense, intel and training. Figure 7.2

is the sum across all of the various group leaders' action choices. There is only 1 attack near the end of the 60 ticks by the populists' militia arm against the Pro Russian Left.

Figure 7 – Russia Scales Up Its Attack Mode as NATO-EU Withdraw



In recent times, Trump and Merkel have both expressed wavering of support for the Baltic nations. To simulate this, we reduced their aid to and security for Latvia, and weakened the model alliances. The agents in the run behave quite differently from the base case. Fig. 7.3 shows considerable increase in economic (dark green) and political (light green) attacks across the entire run. These include Russia constantly attempting to weaken the government, the Latvian economy, and the populists. We aren't showing the charts, but Russia is funneling a lot of aid to the Pro Russian Left and they are likewise participating in the increased turmoil, targeting the government group and the populists. The populists are responding in kind, and with a few more militia strikes against the Pro Russian Left. As Fig 7.4 shows, the military arms of the various groups have scaled back routine things like training, and greatly increased intel collection on each other.

Out of interest to see what the agents would do, we completely removed NATO-EU support in the final run. Given the weakened state of the economy, and the political attacking between groups, Russia swoops in to try and restore order. As Fig 7.5 shows, the economic attacks continue but are accompanied by air and ground forces fighting each other, as well as a small militia response. Fig 7.6 shows the militaries are strictly pursuing kinetic functions only.

StateSim does not simulate governments changing hands, but in this scenario the country is overrun and this is tantamount to a takeover and regime change.

The sum of these runs reveal several things. Politically, the elasticity of conflict with respect to outside support is negative. The more NATO stays involved, the less likelihood of Russian-induced unrest and takeover. An independent SME in the Pentagon was asked to review these runs, and that was his conclusion. He thought the runs reflected what he himself was advising about Russia and the Baltics. From a StateSim agent decision making perspective, these runs also demonstrate how different groups behave toward each other. The Populists and Pro Russian Left groups carry out their values and grievances against each other just as they are profiled. One step ahead thinking works fine for these types of agents/groups. Likewise, the Gray Zone strategic campaign agent (Russia) first used economic power to weaken the government, then turned to use of internal proxy groups to sow unrest, and lastly came in to help restore order (and presumably install a new regime). Multi-step campaign thinking is needed for this type of agent who shifts across a mix of economic, political, informational, and military actions to gain advantage.

5) Conclusions and Lessons Learned

For 20 years, the StateSim research program has asserted that a synthetic approach grounded in systems engineering that combines agent-based (and other) modeling and simulation with best-of-breed reductive social science theories holds the keys to a new path for country modeling and simulation. That new path is the use of descriptive, white box modeling.

It is impossible to conduct true scientific experiments at a macro-historical scale in the real world, so many social scientists tend to use similar cases throughout history as a substitute. The agent society approach is widely cited as a revolutionary new paradigm that permits directly experimenting on a model of the social system of interest. But much of that claim is tied to the notion of illustrating how simple agents and rules lead to emergent phenomena. StateSim moves beyond simple agents to rich cognitive agents built by synthesizing best-of-breed component social science theories. Conducting course of action experiments (what-ifs) with StateSim renders a degree of confidence given its passing so many acceptability or Turing tests over the years as was explained in Section 2.

Throughout the StateSim research program there have been many lessons learned. It is fitting to conclude with a list of the top ten:

1. Science is Not Yet Dead – In this era of ‘big data’, the state of the art for forecasting country instabilities is via statistical extrapolation modeling and pattern fitting. This KISS approach is faster and more accurate than causal modeling as described in Section 1. However, this black box approach is also consistent with the ‘Googlification’ of thinking in the big data and -omics era. One is presumed not to need to understand the social science behind, or the causes of, events (leading to headlines like ‘science is dead’)—one only needs to look up and/or extrapolate the surface of the trends. But this alone is a trap to stop thinking, to accept any trend. As happened with banking during the 2008 recession, those who follow just the metrics of these trending items are vulnerable to bubbles bursting unexpectedly and harmfully. Its unwise to rely on a black box approach that can’t be probed for causation and that one can’t use to explore course of action what-ifs and counterfactuals.
2. Social Science is Naturally Reductive – One wishes to complement trend extrapolation with causal models wherever instabilities seem to warrant a deeper understanding. To do this requires tapping the descriptive knowledge of the social sciences. However, this field is rife with reductive silos containing often-qualitative, narrow best-of-breed theories. It isn’t readily apparent to outsiders how to synthesize these components. Worse still, the diverse social scientists themselves often are content to continue their paradigmatic separatism – economists believe behavior is governed by games with rational actors; human factor modelers believe decision behavior is predicated on micro-processes, cognitive decision speeds, and memory loads; psychologists argue over whether culture vs personality measuring instruments are more significant as indicators of likely behavior; and political scientists often believe that institutions, not individuals, govern behavior. None of the reductive social science niches alone is up to the task as was explored in Section 2. One needs a meaningful synthesis of these different perspectives, and a pragmatic framework to integrate the parts.
3. It is Vital to Have a Unifying Framework – As explained in Section 1 and 2, a Singerian Inquiry framework reliant on decision theoretic mathematics is the approach that StateSim has long used to synthesize reductive components into a larger game of games framework – a sociological game theoretic framework. Indeed, Section 2 of this paper tested H0: The Singerian, expansionistic paradigm of sweeping in best of breed theories across many disciplines will not provide a realistic (valid) recreation of individual countries. Based on 100s of validity and acceptability tests, we rejected H0. The results lead to the conclusion that H1 is accepted: The cross-paradigmatic, synthetic approach leads to valid models. The Singerian approach of sweeping in many theories enriches the agent models and helps users to explain how actions are leading to outcomes and effects.

4. The More KIDS, The Better – An interesting side effect of the synthesis of rich descriptive material (exoteric and esoteric) in the form of diverse social science theories is that the more of it you have, the more it can explain what is going on. From a social constructivist perspective, if the agent approach gets a plausible result (historical recreation, possibility space of futures that indicate the actual outcome, etc.), then this comes with an instant drill down to the causal explanations. This is no longer a black box approach based on correlations alone. Nor is it the simple agent approach. The open box, socio-cognitive, bottom up descriptions entail the “social theory” of what is going on. The richer and more accurate the descriptions, the more likely that design transformation possibilities can be meaningfully studied and explored.
5. Social Scientists Expand Their Theories - Further, when social scientists agree to collaborate under this paradigm they initially tend to pursue their traditional reductive component research and experiments. However, they are then shown the holes and gaps in the agent models to which the implementation of their findings leads. This near-immediate feedback causes them to re-orient their research agenda for the next round of empirical research. The result is expansionism (of their inquiry) and further (better) synthesis with other theories.
6. Country Desk Experts Embellish Their Models - Additionally, when social science professionals (e.g., country and area experts) participate in this process, their qualitative mental models are elicited and quantified through an interview (via the web questionnaire discussed earlier). This confronts them with extrapolations of their models into possible outcomes and allows them to explore causality and mechanism of those effects. Often the experts find they want to refine and embellish their models, especially when they discover omissions and inconsistencies that render the models incapable of recreating historical scenarios or of projecting into the near term in ways that are satisfying to them.
7. Validated Archetype Models Are Reusable – StateSim uses a decision-theoretic mathematics to sweep in best of breed theories from scientists as well as best practice mental models and knowhow of area experts. When we collect these models and validate them via acceptability tests, they are valuable assets. StateSim stores them in a repository and the StateSim Generator figures out how best to reuse them to model each of the other countries. This is possible since there is a large enough collection of reusable types of leaders, followers, and institution archetypes to cover a taxonomy of what is required. As was explained in Section 3, this repository has helped achieve an order of magnitude speedup (from weeks to days) in helping users to create country models.
8. Data Mining Further Speeds Model Construction – One of the core features of StateSim has always been the breakdown of a country into ethno-political groups which dynamically interact with one another within a relevant context. With the help of data mining, the reuse repository, and some simplifying assumptions that adjust the authority of more formal/institutionalized vs. informal/militarized actors based on a country’s level of democracy, StateSim is now able to present a useful baseline picture of a country’s major ethnic, political, and military actors and of the public and private goods and service levels automatically before any user input is needed. This has led to another order of magnitude speedup (from days to hours) to construct the StateSim models as also shown in Sect. 3. It has matured rapidly and effectively, but is still dependent on updates and structured data which is the low hanging fruit. Future research could look into natural language processing to extend these capabilities.
9. Lay Analysts Improve Their Results - when lay intelligence and information analysts seek to use StateSim, it elevates them and acts as a support system that connects them to a wealth of data concerning the region of interest and a repository of pre-built (pre-validated) agent archetypes. It helps them rapidly construct a first approximation model that gives reasonable behaviors and outcomes and results, something they are usually entirely incapable of otherwise creating (not being modelers). Of course, those models can be improved if the analyst-users would consult with country SMEs and/or study the country’s issues at least minimally. For all these reasons the StateSim approach may be seen as a shift to a new methodology and toolset for the analysts, one that helps them drill down into a country or region, exploit expert knowledge about it, explore and test alternative courses of action within it much more rapidly than before.
10. StateSim Research is Unfinished – As a final lesson learned, it should be mentioned that StateSim has matured substantially over the extent of its life, but there will always be a need for continuing development and research. We have succeeded in fielding a number of applications driven by StateSim as was described in Section 2 and StateSim is now usable for modeling developed as well as developing nations. This does not mean that the research program is completed. As the best-of-breed theories embedded inside StateSim evolve, so too must StateSim as an effort to synthesize and apply them. Likewise, the data that is scraped from the cloud comes from varying time periods and doesn’t always yield the latest information on the country being modeled. Until that gets fixed, users need to be cautious and willing to update that manually if needed. Also, since StateSim has been built up over time, some of its components are legacy code and managing to bring all of them along as the newer components are added is an ongoing challenge. Finally, several substantive areas are still being expanded such

as, for example, the economic model (banking sector, monetary supply, trade balance), the strategic gray zone actors' reasoning and campaign steps, and the group demographics and geographic dimensions.

Computational science has altered the nature of each science it has touched to date – physics, chemistry, and biology. It seems inevitable to do the same eventually to the social sciences. There are many ways this could and will happen, but a pathway that the StateSim program illustrates lies in synthesizing the best-of-breed theories from across the narrow, reductive silos that are the social sciences. StateSim also “sweeps in” expertise in the form of reusing past validated models and in mining the cloud for the latest information on the societies of interest. StateSim is fairly unique in pursuing this kind of Singerian inquiry and in doing it successfully, as 100s of validity assessments have borne out. We believe that this type of approach to descriptive, causal modeling is vital for the field, enables course of action explorations, and shines a spotlight on a fruitful way forward.

Acknowledgements

The current project is sponsored by DoD/CTTSO, Australian MoD, and British MoD. Past government sponsors of StateSim include: AFOSR, DARPA, DoD/D9, NIH, ONR, PEO-STRI, US AID, and US Gov. Past private sponsors include gifts from Analog Devices Inc, Anheuser-Busch Foundation, Beck Fund, Boeing, GM Foundation, Lockheed, and projects of numerous students and post-docs over the years. Thanks also to several dozen past collaborators (too numerous to name) over the years.

References

Armstrong, J.S. (2002) Assessing game theory, role playing and unaided judgment. *Inter. J of Forecasting*, 18, 345-352.

Askari, G, Gordji, ME, Park, C (2019) The behavioral model of game theory, *Nature: Palgrave Communications*, 5:57.

Bénabou, R. 2013. “Groupthink.” *Review of Economic Studies* 80: 429-62.

Bharathy, G. and B. Silverman, B (2013), “Holistically Evaluating Agent Based Social Systems Models: A Case Study”, *Simulation Journal: Transactions of the Society for Modeling and Simulation International*. doi: 10.1177/0037549712446854.

Bharathy, G.K. and B.G. Silverman (2012) “Applications of Social Systems Modeling to Political and Business Risk Management” Ch. 17 in Jie Lu, L. Jain and G. Zhang (eds.), *Handbook on Decision Making: Intelligence Methodologies and Applications* Berlin: Springer-Verlag, pp. 331-371.

Boschee, E; Lautenschlager, J; O'Brien, S; et al (2018) "ICEWS Weekly Event Data", Harvard Dataverse, V250, 2018. <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/QI2T9A>

Burns, T. R., Roszkowska, E., Corte, U., & Machado, N. (2017) Sociological Game Theory: Agency, Social Structures and Interaction Processes. *Optimum*, (5 (89)), 187-199. <https://doi.org/10.15290/ose.2017.05.89.13>

Camerer, C, Ho, T-H (2015) “Behavioral Game Theory Experiments and Modeling” Ch. 10 in *Handbook of Game Theory with Economic Applications*, v.4, pp. 517-573.

Cape, M. Lee, H. (2019) “The Athena Simulation: Modeling the Sociocultural Landscape”, Leavenworth: TRADOC G-2.

Center for Systemic Peace (2020) Polity IV Project. <https://www.systemicpeace.org/polityproject.html>

Churchman, CW (1972) *The design of inquiring systems: Basic concepts of systems and organizations*. Basic Books: New York.

Edmonds, B, Moss, S (2005) From KISS to KIDS – an ‘anti-simplistic’ modeling approach. In P. Davidsson et al. (Eds.): Multi Agent Based Simulation 2004. Springer, *Lecture Notes in Artificial Intelligence*, **3415**:130–144.

- Elsaesser, C, et al. (2015) *Computational sociocultural models used for forecasting*, Ch. 10 in *Sociocultural Behavior Sensemaking*, J Egeth (Ed), Washington DC: Mitre Corp.
<https://www.mitre.org/publications/all/sociocultural-behavior-sensemaking-state-of-the-art-in-understanding-the>
- Goldstone, J, Bates, R, Epstein, D, et al. (2010) “A Global Model for Forecasting Political Instability.” *American Journal of Political Science*, 54: 190-208
- Gollin, D (2014) "The Lewis Model: A 60-Year Retrospective". *Journal of Economic Perspectives*. 28 (3): 71–88.
- Halkia, M, Ferri, S, Papazoglou, M, (May 2020) Conflict Event Modelling: Research Experiment and Event Data Limitations, *Proc. AESPEN 2020*, Marseille: LREC, 11–16, p. 42–48.
- Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen et al. (eds.) (2014) World Values Survey: All Rounds - Country-Pooled Datafile Version: <https://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp>.
- Hermann, M. G. (1999) Assessing leadership style. Hilliard, OH: Social Science Automation, Inc.
- House, R.J., Hanges, P.J., Javidan, M, et al. (2004) Culture, leadership, and organizations: The GLOBE study of 62 societies, Thousand Oaks, CA: Sage Publications, 2004
- Janis, I. L., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment*. Free Press.
- Jontz, S (2015) “Data Analytics Programs Help Predict Global Unrest,” in *Signal Magazine*, AFCEA, Dec. 1.
- T.S. Kuhn (1970) *The structure of scientific revolutions*. Chicago: University of Chicago Press
- Lindblom, CE (1959) The Science of “Muddling Through”, *Public Admin Rev*, 19:2, pp. 79-88
- NASA/JPL (2011) “Athena” in *NASA Tech Briefs Magazine*, Washington DC: NASA, Dec. 1.
- Silverman, BG, Bharathy, G. (May 2005), “Modeling the Personality & Cognition of Leaders,” in 14th *Conf on Behavioral Representations In Modeling and Simulation*, SISO (www.sisostds.org).
- Ortony, A, Clore, G.L, Collins, A. (1988) The cognitive structure of emotions. Cambridge: Cambridge University Press.
- Pettersson, T. and M. Öberg (2020) Organized violence, 1989-2019. *Journal of Peace Research* 57(4).
- Silverman, BG, Rees, R., Toth, J, et al. (2005, May), “Athena’s Prism—A Diplomatic Strategy Role Playing Simulation for Generating Ideas and Exploring Alternatives,” in *Proc. of First Internat’l Conf on Intelligence Analysis*, MacLean, VA: Mitre.
- Silverman, BG, Bharathy, GK, Nye, B, Eidelson, (Dec 2007) “Modeling Factions for ‘Effects Based Operations’: Part I – Leader and Follower Behaviors”, *J. Computational & Mathematical Organization Theory*, 13:4, pp. 379 – 406.
- Silverman, BG, Bharathy, GK, Nye, B, Eidelson, (Jun 2008), “Modeling Factions for ‘Effects Based Operations’: Part II – Behavioral Game Theory”, *J. Computational & Mathematical Organization Theory*, 14:2, pp. 120-155.
- Silverman, B. G., Sun, D., Bharathy, G., & Weyer, N., (2016), “Speeding Model Creation Through Reuse: Case of the StateSim Generator,” (Ch. 17 in) J. Cohn, S. Schatz (eds), *Modeling Socio-Cultural Influences on Decision Making*, Boca Raton: CRC Press, , pp. 335-360.

Silverman, B. G., Bharathy, G., & Weyer, N., (2019), “What is a good pattern of life (PoL): Guidance for simulations,” Simulation: Transactions of the Society for Modeling and Simulation International, SMSI, v95, Issue 8,

Silverman, D. 2018. “What Shapes Civilian Beliefs about Violent Events? Experimental Evidence from Pakistan.” *Journal of Conflict Resolution* 63(6): 1460-87.

E.A. Singer (1959) *Experience and Reflection*, Philadelphia: University of Pennsylvania Press.

Solow, R. (1956) A contribution to the theory of economic growth. *Q. J. Econ.* 70, 65–94. doi: 10.2307/1884513

R. Swedberg (2001) Sociology and Game Theory, *J. Theory and Sociology*, 30 (3): 301-335.

USAF (2016) The Effects-Based Operations Approach (EBAO), Montgomery AL: CE LeMay Ctr, ([3-0-D06-OPS-EBAO.pdf](#))

Vogt, M., Bormann, N, Rüegger, S, Cederman, L, Hunziker, P, Girardin, L(2015) “[Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Data Set Family](#).” *Journal of Conflict Resolution* 59(7): 1327–42.