Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.

SIC SEMPER SIMULATION — BALANCING SIMPLICITY AND COMPLEXITY IN MODELING AND ANALYSIS

Ernest H. Page
James R. Thompson

The MITRE Corporation
7525 Colshire Drive
McLean, VA 22102, USA

Matthew Koehler

The MITRE Corporation
55 East Via Plaza Nueva
Santa Fe, NM 87507, USA

ABSTRACT

Determining the level of detail necessary to a modeling effort is fundamental to the discipline. Insufficient detail can limit a model's utility. Likewise, extraneous detail may impact the runtime performance of the model, increase its maintenance burden, impede the model validation process by making the model harder to understand than necessary, or overfit the model to a specific scenario. Intuition suggests that resolving this tension is an intractable challenge that reflects the *art of modeling* and is without promise for general solution. Most analytic communities accept that a truly rigorous, repeatable, engineering solution to the construction of an arbitrary model is unattainable. But the long history of research in modeling methodology suggests there *are* useful steps communities can make in that direction. Through the lens of current modeling challenges, practices and methods in several domains, we hope to add to this important discussion at the intersection of philosophy and engineering.

1 INTRODUCTION

A model is a representation of something else; an abstraction of some other thing or system (commonly denoted as a *referent*). In theory, there can be an infinite range of models for a given referent, each representing a different balance of detail, cost to develop and maintain, runtime performance, sensitivity to underlying assumptions, and myriad other factors. Recognition of this fundamental modeling decision dates to the earliest work in simulation (Conway et al. 1959),

> "Part of the investigator's problem is the selection of the appropriate alternative; he must decide how much of a departure from reality can be tolerated in the interest of simplicity and economy. There are implicit in this procedure the assumptions that the investigator knows enough about the characteristics of interest to intelligently select or construct an alternative system and that he knows enough about the interactions to decide which characteristics of the object system are irrelevant and expendable."

Today, nearly every text in simulation extols the virtues of parsimony in modeling. And compelling examples of effective ensembles of simple models abound (Page 2018). On the other hand ...

> *For want of a nail the shoe was lost.*
> *For want of a shoe the horse was lost.*
> *For want of a horse the rider was lost.*
> *For want of the rider the message was lost.*
> *For want of a message the battle was lost.*

978-1-6654-3311-2/21/$31.00 ©2021 IEEE

For want of a battle the kingdom was lost.
And all for the want of a horseshoe nail.

This well-known proverb, which is said to have been in existence in one form or another since the 14th century, is a compelling portrayal of the nature and consequences of causation. Its spirit and teachings are far-reaching, from parental admonitions to misbehaving children (Franklin 1758) to official opinions expressed by the U.S. Supreme Court (Roberts 2007).

Let us imagine you were hired as a consultant by the Kingdom of Bröpst to develop a kingdom-scale model. You are well-read in simulation practice and theory. You proudly display your autographed copies of (Zeigler, Muzy, and Kofman 2019) and (Law and Kelton 2000). You understand that all models are wrong and some are useful. You have an appreciation for simplicity. Still, the moral of the horseshoe nail proverb seems stark: if you want to reason about the loss of kingdoms, details matter. So you set out with your favorite simulation package, and you begin coding. And coding. And coding. Then one bright, sunny day, you finish. Everything you could think to include about kingdoms, battles, riders, horses and shoes is in there. Your model is impressive.

The nail and shoe performed flawlessly.
For want of oats the horse was lost.

Rats. OK, you add details about the nutritional needs and dietary habits of horses. Being clever, you include variations for several types of horses commonly used by the Bröpst cavalry.

The nail and shoe performed flawlessly.
The well-shod, well-fed horse was on top of its game.
For the want of vertigo medication the rider was lost.

Hmmm. You did not see that one coming. But fair enough, you concede there are non-combat injuries that riders could fall victim to. You code them up. For completeness, you include spread models for all known infectious diseases. Surely, your model is now unassailable.

The nail and shoe performed flawlessly.
The well-shod, well-fed horse was on top of its game.
The rider executed the battle plan without error.
The battle was won.
There was significant collateral damage to the citizenry.
For want of a tax base, the kingdom was lost.

Ah, for crying out loud! Adding representations of micro- and macro-economic forces to your model is beyond the pale. You give up your career as a simulation consultant, and pursue your dream to become the official blogger for your hometown cricket team.

What do we make of this tar pit? The horseshoe nail proverb illustrates that the tiniest detail can have significant impact at the kingdom scale, implying that kingdom-scale models should be highly detailed. But since we cannot include every detail of every situation, what are we left with? We cannot eliminate the possibility that we left out something that is vitally important to the outcome – and it would seem that correctly representing whether or not a kingdom falls would be a pretty big deal. Even if we accept that all models are wrong, some are useful, it would appear that mind-numbing detail is probably required in order to generate *useful* models of kingdom-scale phenomena.

In the remainder of this article, we examine this fundamental tension between detail and simplicity within the current practices in several different modeling domains. We consider the role of complexity science in the creation of models of kingdom-scale phenomena, and discuss differing notions of *usefulness* in modeling and analysis.

2 THEATER-LEVEL DEFENSE SIMULATION

The horseshoe parable has an obvious and direct linkage to defense simulation and analysis. The long history of defense analysis has resulted in a framework for modeling abstraction that is often depicted as a pyramid. The traditional defense M&S pyramid and an extended adaptation by Gallagher et al. (2014) appears in Figure 1. The traditional pyramid conveys a spectrum of modeling detail (also referred to as *resolution*). Engineering level simulations are the most detailed, e.g., representing individual parts and their physics-based interactions on a vehicle. Engagement simulations tend to represent few-on-few interactions. Mission level simulations are typically employed to determine how well a group of systems (sometimes referred to as a force package) work together to accomplish a particular set of objectives. Campaign simulations are the least detailed, used at the policy-making level, and typically represent conflict at a theater scope as force-on-force interactions. The adaptation by (2014) inverts the pyramid to suggest breadth of focus within each level, and adds two additional levels, defense enterprise, and non-military instruments of power.

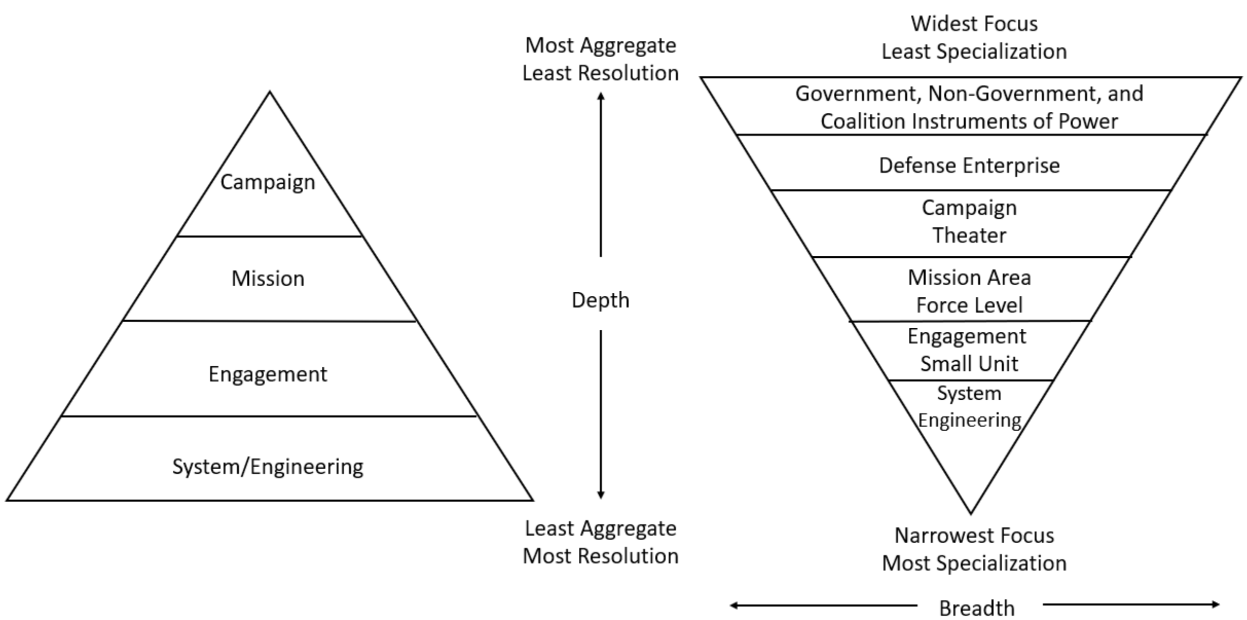

Figure 1: The hierarchy of military models.

In reality, nearly all defense simulations are multi-resolution in nature, with elements that cross the pyramid boundaries. Nonetheless, the levels provide a useful framework for describing prominent loci of analysis conducted within the defense enterprise. In aspiration – if not always in practice – data and results from higher-resolution simulations should be used to inform the lower-resolution simulations, thereby allowing the analysis process to expand in scope while maximizing fidelity without unduly sacrificing runtime performance, complicating maintenance, and so forth. In theory, efficient theater-wide analysis may be "seamlessly" informed by interdependent analyses at the lower (engineering, engagement and mission) levels.

In practice, however, the flow of data and knowledge across the levels of the hierarchy can be kludgy, opaque, expensive and not always embraced by decision makers. Davis (2016) recounts the events and rationale behind the loss of support for campaign modeling within the U.S. defense policy and strategy making process circa 2011. One of the primary criticisms of campaign analysis involved the opacity of the underlying models and simulations. Decision makers expressed frustration at the volume of input parameters, layers of aggregation, litany of assumptions, and the lengths to which analysts had to go to explain the results. Akst (2013) sympathizes with the decision maker pushback on the analysis, and suggests that decision maker frustration was partly due to unrealistic expectations with respect to the utility of campaign simulations:

> "[Campaign simulations] can support the examination of alternative force structures or systems in their ability to achieve the mission. They can provide good insight into such things as the sequencing of force arrivals or consumption of various supplies and munitions. They can even help a commander explore the relative merits and shortfalls of alternative concepts of operation. However, we get into trouble when we start treating the outcomes of these simulations as predictive. I would caution against using campaign simulations to predict equipment losses, time to achieve objectives, or casualties."

Given the current state of campaign simulations, Akst provides sound, practical advice. But his position raises the question: are predictive models of campaign scale equipment losses, time to achieve objectives, and casualties an impossibility? Is *any* meaningful level of prediction supportable? Can we quantify the uncertainties and rationally bound our predictions? Effective policy making and investment planning would clearly benefit from such capabilities. Ongoing research in deep uncertainty may offer some help here (Bankes 1993; Kwakkel and Pruyt 2013; Tolk, Dinh, Comer, and Scott 2020).

Similarly, can we do better in understanding which details matter and in what contexts? Meyer et al. (2005) describe experiments involving the relationship between probability of kill (Pk) and terrain resolution in engagements with point and area munitions. They show that when the terrain resolution is low, high resolution estimates of Pk lead to overestimates of weapon effectiveness. In network simulation, Abdelhafez and Riley (2006) show that commonly adopted abstractions of certain protocols provide execution efficiency without unduly sacrificing accuracy when network traffic is low, but these abstractions lead to overestimates of network performance by as much as 50% under higher loads. It seems reasonable to believe there are similar risks to campaign analysis. Can we identify critical elements that are typically over-abstracted at the campaign level?

A recent Defense Science Board (DSB) study (2021) suggests a dual research and investment strategy for campaign-level analysis: one path geared toward the development of simple, explainable models à la (Page 2018), and a complementary path to reinvigorate traditional campaign modeling such that the linkages to higher resolution analysis are more transparent, explicit, and effective.

The DSB findings, and Akst's challenge to the defense analysis community are a call to reevaluate fundamental approaches to model conceptualization and implementation, and to challenge the foundational assumptions in our approach to defense analysis. Warfighting technologies and strategies evolve, and the geospatial extent of the warfight expands dramatically, while the commander's decision and reaction time shrinks to the threshold of human cognition. Our approaches to defense analysis must evolve to meet these challenges. There seem like a lot more horseshoe nails to deal with these days. There is a pressing need for reliable, quantifiably predictive, analysis of theater (and above) scale phenomena.

3 IMPLICATIONS AND CONTRIBUTIONS FROM COMPLEXITY SCIENCE

The fundamental tension between detail and simplicity is inherently tied to the concepts of complexity science. In the late 1940's Warren Weaver (1991) laid out three broad periods of scientific advancement covering roughly three centuries. And there are fascinating parallels in Weaver's outline with the history of simulation laid out by Goldsman, Nance and Wilson (2010).

3.1 Problems of Simplicity

From the seventeenth century to the late nineteenth century, Weaver claims that science was restricted to what he deemed *problems of simplicity*. These problems often involved the interaction between only two variables, while everything else was held constant. This period includes Buffon's needle experiment, where he estimated π using Monte Carlo simulation techniques (Goldsman, Nance, and Wilson 2010). Buffon's random number generator was just a needle tossed onto a table, but using the geometry of lines drawn on the table and the length of the needle, he boiled the problem down to two variables, the probability his needle touched one of the lines and the constant π he wanted to measure. Arguably, problems of simplicity make up the foundations of the hierarchy of defense models. Physics-based interactions constitute the most notable scientific work of this period, and they are key components of many larger scale simulations.

3.2 Disorganized Complexity

Buffon's experiment came toward the end of Weaver's first period of problems of simplicity and it foreshadows the next evolutionary step in science and simulation. When more than two variables begin to interact the techniques refined for problems of simplicity are no longer tractable. Confronting this type of problem gives rise to Weaver's second period, *disorganized complexity*. Disorganized complexity is dominated by the techniques of probability and statistics. Although Weaver does not mention it directly, those versed in simulation will recognize the out-sized role the central limit theorem plays in disorganized complexity. Problems of any number of variables (such as the billions of molecules in an ideal gas) are abstracted into the statistical average. Here again, Weaver's characterization fits neatly into the history of simulation, overlapping the work of William Sealy Gosset (Student) and his derivation of the t-distribution in 1908 (Goldsman, Nance, and Wilson 2010) and bringing us to the mid 1900s. Once intractable mathematics were now abstracted into draws from a probability distribution and population level conclusions could be reached. These approaches jump over a few levels in the defense simulation hierarchy to the campaign level and above, as details of individual skirmishes are lost in the statistical average of the entire force.

3.3 Organized Complexity

The Achilles heel of the central limit theorem is the assumption that the variables in question are, at least approximately, independent and identically distributed (i.i.d.). When this assumption holds true, Gauss's Normal distribution, with its symmetry and easy math, will emerge to rescue us from complexity. Combined with the law of large numbers, we need only crank up the replications to achieve a steady-state and broad answers to complex problems can be obtained. But when the assumptions of i.i.d. are no longer valid, we enter Weaver's final period of *organized complexity*, borne out of World War II and continuing to the present day. Organized complexity is characterized by increasingly dependent, often exceedingly diverse, variables altering their rules for interaction dynamically. Such systems rarely reach a steady-state and are often characterized by mathematical chaos. They are prone to unpredictable cascades, long memory in their empirical auto-correlation functions, and heavy-tails (i.e., infinite moments) in the empirical probability distributions derived from their behavior. But they also often arise from a simple set of rules at the finest scale. Schelling's model of segregation (C.Schelling 1969) or Reynolds model for flocking behavior (W.Reynolds 1987) are based on remarkably simple automata following relatively simple rules. When they are scaled to a suitably large number of agents, however, patterns of behavior emerge that would be unexpected if examining only one or two such agents through a strict reductionist approach and then extrapolating the average behavior to the population. How could one understand an ant colony by studying an ant in isolation, but how many ants are required to make a colony?

In 1948, Weaver was hopeful these problems would be solved using two recent advancements from World War II; computers and diverse teams. As simulation practitioners, we are all familiar with the value computers bring to scientific advancement. But it is the field of complexity science that has fully embraced the idea of diverse teams tackling a single problem. In order to connect the steps in defense simulation hierarchy, the concepts of physics need to be combined with military doctrine, intelligence, computer science, foreign policy, political science, operations research, psychology, and sociology at a minimum. The necessity of diverse teams is inescapable.

3.4 Complexity Science - A Border Modeling Example

Today we tend to think of organized complexity as synonymous with "complexity science" and the study of complex systems. A useful definition of complex systems is "systems that comprise many interacting parts with the ability to generate a new quality of macroscopic collective behavior through self-organization" (Sornette 2006). An example of such a system encompasses the ongoing crises on the United States' border with Mexico. The U.S. is currently engaged in heated debate about how to handle the constant influx of unauthorized immigrants on the southern border. The components of this system include the unauthorized immigrants, Customs and Border Protection and the U.S. Border Patrol, surveillance technology, impedance technology, politicians, voters, and the politics, economics, and crime rates of the foreign nations many of the immigrants are fleeing. We can deconstruct this complex system into more manageable pieces and ask simpler questions, such as what is the best configuration of manpower and surveillance technology to maximize U.S. Border Patrol's effectiveness? But even this simple question requires us to define a measure of effectiveness, quantify the unauthorized immigration rate, consider all possible configurations of surveillance and manpower, and estimate the relationship between manpower and technology. It is immediately obvious that non-linear, complex relationships exist. More surveillance generally leads to more detections of unauthorized entries. But more detections will also increase the utilization of each Border Patrol Agent. Individual Border Patrol Agents have finite capacity and will ultimately be too busy with current detections to respond to new ones. Similarly, more apprehensions will lead the unauthorized immigrants to adapt their behavior in an attempt to avoid being caught. Success in one sector could therefore be immediately offset by increased flow into an adjacent sector.

This system has components that span the defense simulation hierarchy and is the poster-child for organized complexity. Moreover, it has been successfully explored with an agent-based simulation that incorporates all three aspects of Weaver's historical progression. The surveillance technology and motion of the human agents in the model boil down to straightforward physics (problems of simplicity). The arrival rate and location of arrivals can be accurately depicted with draws from a compound Poisson random variable that considers each arrival as a group of unauthorized immigrants, where the group size is also a random variable (disorganized complexity). And finally, these components are combined with models of diverse human behavior that includes law enforcement agents, peaceful economic migrants seeking asylum, and criminal elements with more nefarious intentions (organized complexity). The key component to the success of this model was a diverse team that included people with direct border experience, computer scientists, operations researchers, cartographers, physicists, and industrial engineers. Equally important was a well-defined objective that hinged on the marginal difference between end states under controlled configurations of manpower and technology. The net result was a model that accurately quantified the relationship between manpower and technology, allowing a $15 billion per year organization to make informed decisions about their investments (Thompson and Rosen 2018).

3.5 Complexity Science - A Pandemic Modeling Example

The model described above would likely satisfy Akst's desire to limit simulations to quantifying the contribution of certain assets to the warfight. But it would be short-sighted to claim that such quantification is unrelated to prediction. One of the most valuable outcomes of simulation modeling is knowledge gained by the modelers. They inevitably come to know their own system in ways they were never forced to contemplate before. *Those insights lead to predictive abilities.* A case in point is an agent-based model of the recent COVID-19 pandemic. In the United States, state and local authorities were under extreme pressure to decide between keeping the citizenry safe with non-pharmaceutical interventions (NPIs) that crushed the economy or risking wide-spread outbreaks and deaths while allowing the economy to thrive. Shifting the decision-making to the local authorities illuminated a well-known problem with standard differential equation models of disease propagation; namely homogeneous mixing. Will a person in New York City have the same probability of an infectious contact as a person in Cheyenne, Wyoming? The obvious answer is no, but the more one analyzes this problem, the deeper the rabbit hole goes. Even if we restrict the model to a given location, are all age groups equally likely to interact or are people of all occupations equally likely to interact? And the more we subset the population, the more we slowly degrade the simplifying assumptions that make differential equation models tractable. Once again, however, complexity science offers some relief. By shifting to an agent-based approach that starts with a network structure, the individual probabilities of contact are captured by the network's edges and the decisions of each agent in the model. The impact of NPIs can be modeled as a dynamic change to that network structure that some agents will adhere to while others rebel against. A simulation of this type categorizes which counties in the state of Maryland were at greater risk of outbreak than others if NPIs were lifted too soon (Koehler, Slater, Jacyna, and Thompson 2021). Interestingly, while the authors of this model were pursuing publication, the state of Maryland decided to reopen. A subset of the 24 counties in Maryland refused to follow this reopening schedule due to their case loads at that time (Nirappil, Cox, and Wiggins 2020). These counties corresponded exactly to the counties categorized as being at highest risk according to the model. This outcome suggests that prediction and actionable information can be reliably gleaned from complex simulation models if they intelligently combine the problems of simplicity with disorganized and organized complexity. Achieving this capability within the defense simulation hierarchy may result from combining our ever-advancing computing power with all the tools we have obtained from simplicity to complexity and mixing it altogether with diverse teams consisting of strange bedfellows.

4 PREDICTION AND EXPLANATION

Our need for predictive models is widely accepted. But what about *explanatory* models? While prediction and explanation may not be orthogonal, they are also not identical. One can fit a complicated function to a time series and do a very good job of predicting the next few values of said time series. But, that feat does nothing to explain why the time series is developing in a particular way. On the other hand, we can do a very good job of explaining how something works while providing little predictive capability. The theory of evolution, for example, explains why species change over time but does not provide us with a means to predict species change over time, other than the fact that it will occur under the right conditions.

Should we prefer explanation or prediction? That depends on the question, or at least it should. The process is more, or less, difficult based on where the problem fits within Weaver's characterization and on the needs driving the modeling effort. "Well behaved" systems such as Weaver's simple or disorganized complex systems can be easier to predict and explain. When one deals with organized complex systems, however, prediction and explanation can be much more difficult. Under these circumstances one must carefully define the modeling effort being undertaken or it can very quickly grow out of control otherwise. In line with Figure 1, is the simulation being used as a: (a) Thought experiment, problem definition or basic requirements analysis? (b) Functional analysis, or to experiment with basic generating mechanisms or coarse-grained forecasting? Or, (c) Synthesis of components, physical validation or fine-grained forecasting? Finally, is there a referent that can be used to define or understand how well the simulation represents the phenomena of interest?

On the coarser end of the uses, (a) above, one likely requires only relational equivalence to a referent (Axtell, Axelrod, Epstein, and Cohen 1996) if there is one. Moreover, at this end of the spectrum, one likely only needs Level 0 or perhaps Level 1 Empirical Relevance (Axtell 2005). Here the components of the simulation only need to behave plausibly (qualitative micro-level relevance (level 0) or qualitative macro-level relevance (level 1)). As the needs and use of the simulation become more specific, e.g., some form of forecasting or one begins to ask specific questions about the system's generating mechanisms, then the simulation should more closely align with the referent. Here distributional equivalence/level 2 empirical relevance is the goal, specifically distributional agreement, i.e., the aggregated distributions produced by the simulation and referent are statistically indistinguishable, between the simulation and referent (Axtell 2005; Axtell, Axelrod, Epstein, and Cohen 1996). Finally, as the needs of the analysis require fine-grained forecasting or ask very specific questions about the generating mechanisms of the system, one strives for identical output between the simulation and the referent driven by micro-level quantitative correspondence (Axtell 2005; Axtell, Axelrod, Epstein, and Cohen 1996). While truly identical output between the referent and simulation is unlikely, it would be ideal. Moreover, the data requirements for micro-level quantitative correspondence are very high and may be very difficult to obtain, though given the data collection going on today, this bar may be getting lower.

In many ways the above outlines a process to define what an adequate simulation is within a given context. Pulling that thread further, one could then use it as a way to define the necessary validation process is for the simulation in question. While V&V can easily become a box checking exercise, especially in resource constrained situations. There is a wealth of guidance on tailoring V&V processes (Balci 1998). Is the simulation being used to create engineering-grade requirements or specifications? Or is the simulation being used to inform a thought experiment? The intended use drives the validation exercise. If the simulation is simply a tool aiding a thought experiment, the necessary validation may be just that the components behave in plausible ways and interact in reasonable ways. If, on the other hand, one is deriving specific engineering requirements from the simulation, then the validation exercise must be much more rigorous.

A framework for V&V of agent-directed simulations within a systems engineering context can be found in (Barry, Koehler, and Tivnan 2009). Beyond the validation question, this framework can also be used at the beginning of a modeling endeavor to define how far down the horseshoe nail road one should go down and to define success (validation) so you know when you get there. This structured approach also helps mitigate potential artificialities of the level-resolution demarcation in Figure 1. To many, the figure implies that as the scale of the situation being simulated increases the level of resolution must uniformly decrease. As noted in Section 2, in practice most simulations have elements that cross resolution boundaries. For example, if a campaign-scale question hinges critically on communications then a simulation may include a relatively low resolution representation of forces with high resolution communications. Additional considerations for these trades are discussed in (Meyer, Koehler, Barry, and Tivnan 2005).

By thinking critically about the question, the "thing" to be modeled, the necessary insights and purpose, and the current understanding of how the "thing" works, one can use the framework outlined above to define what should be in and out of the model and what success looks like so you know when you are done. In this sense you can figure out whether or not you need a well-fed horse or simply one with feet that could be shoeless.

5 CONCLUSIONS

In this article we explore the balance between simplicity and complexity in simulation modeling. Considered through the lens of defense simulation and analysis, the balance must reconcile the need for detailed physics-based representations of components at the system engineering level with the 'softer' machinations of global politics at the enterprise and government levels. This tension aligns well with the history of complexity science, as interpreted by Weaver's progression from problems of simplicity that focus on just a couple of variables, to disorganized complexity that exploit statistical averages and the central limit theorem, to organized complexity that relies heavily on computer simulation and diverse teams to uncover the emergent effects. Ultimately, the balance between simplicity and complexity is found in the original objective of the modeling endeavor. The seemingly subtle distinction between prediction and explanation is accentuated by organized complex systems, where explaining the mechanism often offers little predictive power and vice versa.

The common thread through these dizzying topics is that the work surrounding the construction of a simulation is as important – or perhaps more important – than the resulting simulation itself. It has been said, in a number of different ways, that effective analysis is a combination of both analytical rigor and creative insightful thinking (Rumelt 2012), (Betts 2000), (Courtney, Kirkland, and Viguerie 1997). However, that does not mean that a general solution is out of reach. Indeed, the simulation literature provides a framework that can guide simulation modelers through the journey from simplicity to complexity to generate the appropriate model for the problem in question. One must first determine the boundaries of the system and the objectives they seek. Is the goal to predict or explain or explore or train? Will the real-world decisions be made at one scale or across multiple scales? Are all the details important or will a marginal comparison suffice? Once the objective is defined, it should give way to testable hypotheses which in turn suggest the appropriate model and level of detail required.

Creating a *useful* simulation is difficult. Knowing what features should be added and at what level of resolution is critical for creating the insights necessary to answer the question established by the objective. All is not lost. There are general guiding principles and domain-specific practices, some of which we have addressed here. The challenge to simulation modelers is to assemble the appropriate diverse team, tailor their thinking based on the progression from simplicity to organized complexity, and remember that the most valuable result of simulation is often the insights gained by the modeling team.

In this article, we admit raising more questions than solutions offered. We hope our assessment of the fundamental tension between detail and simplicity as reflected within current practices in several different modeling domains, and framed by complexity science, contributes to this important topic in modeling methodology.

ACKNOWLEDGMENTS

The opinions expressed in this work are those of the authors, and do not represent official positions of The MITRE Corporation or the U.S. Government. The authors thank Andreas Tolk for the many conversations that shaped shaped this work, and the referees for their valuable feedback on early drafts. Ernie would like to thank Dick Nance for a priceless education in technical writing, and Barry Nelson for his Keynote address at WSC 2017, and reminders that it is OK to have fun at work.

REFERENCES

Abdelhafez, M., and G. F. Riley. 2006. "Revisiting Effects of Detail in Ethernet Simulations". *SIMULATION SERIES* 38(2):297.

Akst, G. 2013. "Campaign Modeling: The Fat Lady Has Not Yet Sung". *Phalanx* 46(1):5–6.

Axtell, R. 2005. "Three Distinct Kinds of Empirically-Relevant Agent-Based Models". Technical report, Brookings Institution.

Axtell, R., R. Axelrod, J. Epstein, and M. Cohen. 1996. "Aligning simulation models: A case study and results". *Computational and Mathematical Organization Theory* 1(2):123–141.

Balci, O. 1998. "Verification, validation, and testing". In *Handbook of Simulation*, edited by J. Banks, Chapter 10, 335–393. John Wiley and sons.

Bankes, S. 1993. "Exploratory modeling for policy analysis". *Operations Research* 41(3):435–449.

Barry, P., M. Koehler, and B. Tivnan. 2009. "Agent-Directed Simulation for Systems Engineering". In *Agent-Directed Simulation and Systems Engineering*. John Wiley Publishing.

Betts, R. K. 2000. "Is strategy an illusion?". *International security* 25(2):5–50.

Conway, R. W., B. M. Maxwell, and W. L. Johnson. 1959. "Some Problems of Digital Simulation". *Management Science* 6(1):92–110.

Courtney, H., J. Kirkland, and P. Viguerie. 1997. "Strategy under uncertainty". *Harvard business review* 75(6):67–79.

C.Schelling, T. 1969. "Models of segregation". *The American Economic Review* 59(2):488–493.

Davis, P. K. 2016. "Capabilities for Joint Analysis in the Department of Defense". Technical report, RAND.

Defense Science Board 2021, January. "Final Report of the DSB Task Force on Gaming, Exercising, Modeling and Simulation". Technical report, Defense Science Board, Washington, D.C.

Franklin, B. 1758. *Poor Richard's Almanac*. Facsimile edition ed. 2. The Complete Poor Richard Almanacs.

Gallagher, M. A., D. J. Caswell, B. Hanlon, and J. M. Hill. 2014. "Rethinking the Hierarchy of Analytic Models and Simulations for Conflicts". *Military Operations Research* 19(4):15–24.

Goldsman, D., R. E. Nance, and J. R. Wilson. 2010. "A brief history of simulation revisited". In *Proceedings of the 2010 Winter Simulation Conference*, 567–574. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Koehler, M., D. Slater, G. Jacyna, and J. R. Thompson. 2021. "Modeling COVID-19 for Lifting Non-Pharmaceutical Interventions". *Journal of Artificial Societies and Social Simulation* 24(2).

Kwakkel, J. H., and E. Pruyt. 2013. "Exploratory Modeling and Analysis, an approach for model-based foresight under deep uncertainty". *Technological Forecasting and Social Change* 80(3):419–431.

Law, A. M., and W. D. Kelton. 2000. *Simulation modeling and analysis*. third ed. McGraw-Hill New York.

Meyer, T., M. Koehler, P. Barry, and B. Tivnan. 2005. “How Simple Is Simple Enough? Military Modeling Case Studies”. In *Proceedings of the Agent 2005 Conference*. Lemont, IL: Argonne National Lab.

Nirappil, F., E. Cox, and O. Wiggins. 2020. “Hogan announces gradual reopening; D.C. shutdown extended through June 8”. The Washington Post. Accessed May 13 2020. https://www.washingtonpost.com/local/dc-extends-shutdown-maryland-to-reopen/2020/05/13/370c6fb6-951f-11ea-82b4-c8db161ff6e5_story.html.

Page, S. E. 2018. *The Model Thinker*. New York, NY: Basic Books.

Roberts, J. G. 2007. “*Massachusetts v. Environmental Protection Agency*”. 549 U.S. 497 (dissenting opinion). Accessed May 11 2021. http://www.law.cornell.edu/supct/html/05-1120.ZD.html.

Rumelt, R. P. 2012. “Good strategy/bad strategy: The difference and why it matters”. *Strategic Direction* 28(8).

Sornette, D. 2006. *Critical phenomena in natural sciences: chaos, fractals, self organization and disorder: concepts and tools*. Springer Science & Business Media.

Thompson, J. R., and S. Rosen. 2018. “Internal Verification and Validation of the Quantitative Decision Making Model”. Technical report (limited distribution), The MITRE Corporation.

Tolk, A., K. Dinh, K. Comer, and S. Scott. 2020. “Exploratory Analysis to Address Deep Uncertainty–Using Calibratable System Models for Exploratory Simulation of Complex Missions”. In *2020 Spring Simulation Conference (SpringSim)*. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Weaver, W. 1991. “Science and complexity”. In *Facets of systems science*, 449–456. Springer.

W.Reynolds, C. 1987. “Flocks, herds and schools: A distributed behavioral model”. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, 25–34. Association for Computing Machinery.

Zeigler, B. P., A. Muzy, and E. Kofman. 2019. *Theory of Modeling and Simulation: Discrete Event and Iterative System Computational Foundations*. third ed. London, UK: Academic Press.

AUTHOR BIOGRAPHIES

ERNEST H. PAGE is Chief Engineer of the Office of Secretary of Defense (OSD) Research and Advanced Capabilities Department at The MITRE Corporation. He received the Ph.D. in Computer Science from Virginia Tech. He holds a variety of service and editorial roles within the simulation community, and was a member of the Winter Simulation Conference Board of Directors from 2001-2012. During 2019-2020, he served on the Defense Science Board (DSB) Task Force on Gaming, Exercising, Modeling and Simulation. His email address is epage@mitre.org.

JAMES R. THOMPSON is an Operations Research Principle at The MITRE Corporation. He holds a BS in Industrial and Systems Engineering from Georgia Tech and an MS and PhD in the same field from North Carolina State. His email address is jrthompson@mitre.org.

MATTHEW KOEHLER is the Computational Finance Group Lead at The MITRE Corporation. He received the JD at George Washington University Law School and the Ph.D. in Computational Social Science at the George Washington University. He currently serves as the President of the Computational Social Science Society of the Americas. His email address is mkoehler@mitre.org.