

Lessons Unlearned: Barrier or Key to the Future?

S. K. Numrich
Institute for Defense Analyses
Alexandria, VA
Snumrich@ida.org

ABSTRACT

Those of us who work for the military in some capacity are well aware of the emphasis placed on lessons learned. Army's recognition of the importance of learning from past experience is embodied in the Center for Army Lessons Learned (CALL). CALL *identifies, collects, analyzes, disseminates, and archives lessons and best practices ... to facilitate the Army's ... adaptation to win wars.* There is great wisdom in the practice of reflecting on our experiences for building a better future in a complex world. When we truly learn a lesson, we incorporate it into our practices to advance our knowledge and capability, and to improve our simulation products. But what of lessons unlearned, those things we have tripped over, documented, forgotten and thus have tripped over again. Is there a role for them? What about our failures, the ones we hesitate to celebrate in papers and presentations? Are we neglecting a valuable resource? Enter the notion of superforecasting. In 2010, the Intelligence Advanced Research Projects Agency (IARPA) issued a Broad Agency Announcement (BAA) entitled Aggregative Contingent Estimation (ACE) with the goal of dramatically enhancing the accuracy, precision and timeliness of intelligence forecasts for a wide range of event types. Among the participants, a newly developed program, the Good Judgment Project (GJP), aimed at harvesting the "wisdom of the crowd" while simultaneously examining the performance of participating individuals. About 2% of the 250 individuals in the "crowd" emerged as superforecasters who beat the benchmarks by as much as 30%. That result would be of little interest, except that superforecasting capability can be trained. The thrust of this paper is an examination of how the thought patterns for superforecasters could influence how we work as program managers, technologists and trainers to improve our products and perhaps contribute to training more effective, agile military leaders. And, yes, unlearned lessons are telltale symptoms of not thinking like superforecasters. But imagine where we could take our industry if we could improve by only 10% our ability to make better judgments and assess more accurately potential futures.

ABOUT THE AUTHORS

Susan K. Numrich (Sue) PhD, CMSP, has contributed to the science and technology of Modeling and Simulation for over 50 years. She began her career in physics at the Naval Research Lab (NRL) working at the engineering level of modeling and simulation. She was fortunate to have support from NRL to pursue graduate work at The Johns Hopkins University, American University and Cambridge University (UK). She was selected by NRL and its parent organization, the Office of Naval Research, as their representative to the Navy's Modeling and Simulation Management Office. When The Technical Cooperation Program (US, UK, CA, AUS, NZ) decided to explore distributed simulation as an area of international interest, Sue was selected to lead the development, an effort for which she received both Project and Individual Performance Awards. She was subsequently selected to serve on NATO's newly restructured Studies, Analysis and Simulation Group. Her last three years as a civil servant, Sue served as the Director of Technology for the Defense Modeling and Simulation Office where she was exposed to simulation across the whole of the Department of Defense. Since 2005, she has been a research staff member at the Institute for Defense Analyses where she has contributed to studies in the use of military simulation, the incorporation of human activity and behavior into various types and levels of simulation, and the validation of a variety of simulations. Initially a member of Simulation and ECIT subcommittees, Sue founded and was the first chair of the Tutorial Board, now a staple of IITSEC week, offering introductory through advanced tutorials on a wide variety of relevant topics to help educate the IITSEC community. Along the way, Sue was awarded the Office of the Secretary of Defense Exceptional Civilian Service Award, received a patent, became a Fellow of the Acoustical Society of America, has written four book chapters, over 50 technical papers and has two Academic appointments spanning fifty years.

Lessons Unlearned: Barrier or Key to the Future?

S. K. Numrich
Institute for Defense Analyses
Alexandria, VA
Snumrich@ida.org

INTRODUCTION

This paper begins by briefly looking at the importance of forecasting and then by describing the ground-breaking work by Peter Tetlock and Dan Gardner as documented in their book, *Superforecasting: The Art and Science of Prediction*. (Tetlock and Gardner, 2015). Tetlock contends that foresight is not a gift, but rather a product of a particular way of thinking; therefore, the paper will explore some of the characteristics of how superforecasters think. Once these patterns have been unearthed, the paper will show how such patterns could be used to improve our processes and perhaps change our minds about how we treat success, failure and lessons unlearned.

NO BETTER THAN “CHIMPANZEES THROWING DARTS”

Why should we be interested in forecasting? To begin with, we are all practitioners of the art. At first blush, that sounds very strange and not at all like us, but if we reframe the statement by stating that all of us assess the future along many dimensions in our daily lives and make decisions based on that assessment, we might just reconsider.

We assess the climate of the future when we decide to buy a house, invest in ALPHABET (very fortunate if you did some years ago), change jobs, opt for another degree, buy a new high-tech home security system, among other decisions. We routinely depend upon forecasts, weather being perhaps the most ubiquitous. In fact, meteorologists are really quite good at forecasting a day, a week, sometimes even ten days in advance, but ask them about the weather in three months and they revert quickly to reams of climatology data – not forecasts. So we know that there are time limits to how far we can see reliably into the future.

The technologists among us are constantly betting on the future when we choose the technologies we employ. We do it, too, when we select a Galaxy over an iPhone or a Chromebook over Surface. And we don't always choose well. Remember when we chose the BETA recorder for its better technology only to have the future captured by VHS? We humans are notoriously poor at predicting the future, and like the meteorologist, the farther out into the future we go, the worse we get. But how bad is bad?

That was a question Daniel Kahneman posed to colleague Philip Tetlock – a question that kicked off twenty years of measuring the track record of experts across a wide variety of fields including economy, stocks, elections wars and other issues of the day. In the end, the average expert did as well as random guessing. The research was documented (Tetlock, 2005) and was the most extensive, evidence-based assessment of expert judgment in scientific literature at that time. However, much to his current regret, Tetlock characterized the randomness as a chimpanzee throwing darts – a far more catchy phrase than random guessing and one that has stayed with the study. So, we're not very good at making predictive judgments.

Resurgence of Collective Intelligence

We know a topic is important (or exciting) when a dramatic television series is shaped about the topic. Such is the contemporary embodiment of collective intelligence – the wisdom of the crowd. The notion of collective intelligence is far from new. In the opening of his book on *The Wisdom of Crowds* (Surowiecki, 2004) Surowiecki recounts an experiment done by Francis Galton (Galton, 1907), a Victorian era statistician, in which attendees at an English country fair attempted to guess the weight of a prize ox. Galton sorted the guesses and found the midpoint as the *vox populi* or voice of the people. The value of the weight at this midpoint was 0.8% of the actual weight of the ox.

Collective intelligence has been around for millennia, according to members of the Collective Intelligence program at Massachusetts Institute of Technology (MIT), at least insofar as groups of people have come together and functioned

as an intelligent unit. However, more recently, the notion of collective intelligence has been applied to specific types of collections of individuals working in a systematic pattern.

The Delphi method attempts to refine the process of group deliberation by involving a panel of carefully selected, independent experts answering questions in two or more rounds. After each round a facilitator provides a summary of results with reasons for the judgments. Answers are revised in light of others' replies with the expectation that the responses will converge to the "correct" answer.

Human swarming was designed as an optimized method for unleashing the wisdom of crowds. The method often uses individuals logged into a platform that receives and publishes responses allowing groups of distributed users to collectively answer questions, generate ideas and make predictions as a singular emergent entity. Early testing shows that human swarms can out-predict individuals across a variety of real-world issues.

Enter IARPA and the ACE Project

With the resurgence of the notion of collective intelligence, IARPA launched an effort to measure collective intelligence's ability to assist intelligence analysts in their work. In 2010, the IARPA (IARPA, 2010) issued a Broad Agency Announcement (BAA) entitled Aggregative Contingent Estimation (ACE) with the goal of dramatically enhancing the accuracy, precision and timeliness of intelligence forecasts for a wide range of event types.

IARPA choose to execute the program as a tournament in which teams of researchers competed against each other and an independent control group. Margins of success were extremely high – requiring that teams beat the control group by as much as 20% the first year and 50% by the end of the program. Teams could be recruited, trained and run in any way that research group chose, opening the process up to almost limitless possibilities for innovation.

Among the participants a newly developed program, the Good Judgment Project (GJP) aimed at harvesting the "wisdom of the crowd" while simultaneously examining the performance of participating individuals. Tetlock's team was composed of volunteers from every walk of life, with all sorts of training and experience. Each year he kept the best and recruited more, continually keeping the best and weighting the forecasts in favor of those who revised their forecasts more often and turned in the best results. The unanticipated discovery was that there were ordinary people who continually outperformed the rest and did so by significant margins – the superforecasters.

Who were these superforecasters? They were not the credentialed and acclaimed experts. One was a retired pipe installer, another a former ballroom dancer. The good news for all of us is that the superforecasters don't have special genetics or uncommon luck. In a research paper published subsequent to Tetlock's team winning the IARPA tournament by wide margin, Mellers (Mellers, 2015) found four mutually reinforcing explanations of superforecaster performance:

1. Cognitive abilities and styles
2. Task-specific skills
3. Motivations and commitments
4. Enriched environments

These four factors suggest that superforecasters are partly discovered and partly created. That leads to the question of how they can be created and what natural traits are required in a trainable individual.

SUPERFORECASTERS: THOUGHT PATTERNS AND PROCESSES

Superforecasters are somewhat above average intelligence, but by no means geniuses. Their abilities derive from a natural curiosity and desire to learn, coupled with particular ways of thinking, of gathering information, of updating beliefs. These habits of thinking, according to Tetlock, can be learned and cultivated by any intelligent, thoughtful, determined person. The characteristics of the superforecaster and the specific patterns of thought are the subject for the rest of this section. Because we as a community seek to equip warfighters with the ability to plan and make decisions involving the future, look toward the evolution and emergence of technology to support our tools, and work to rationalize our processes encouraging the best decision making from requirements to deployment – we need to absorb these patterns of thought across our members. According to the *Economist* (Economist, 2015), the techniques

and habits of mind set out by Tetlock are a gift to anyone who has to think about what the future might bring – in other words, to everyone.

The Right Blend of “Thinking Fast and Slow”

Nobel Prize winner (Economics, 2002) Daniel Kahneman, noted for his work on the psychology of judgment and decision making, advanced the notion of dual process reasoning in his book *Thinking, Fast and Slow* (Kahneman, 2011). According to this notion, we use two modes of decision making: System 1 (thinking fast) is automatic and low effort, tending to favor narratively coherent stories over careful assessments of evidence, while System 2 (thinking slow) is deliberate, effortful and focused on logically and statistically coherent analysis of evidence.

We tend to rely on System 1 most of the time and that’s appropriate for the myriad daily decisions we make. Were we to rely on the laborious, time-consuming System 2, we would fall into the effect often known as “analysis paralysis,” in which we constantly seek more evidence and become progressively less willing to make the essential decision. Our mental rules of thumb (System 1), the heuristics by which we make so many decisions, are systematically biased and often in strange ways. Our experience becomes our touchstone bringing with it the attitude that our personal experience is the only relevant evidence for the decision at hand. Our probability estimates are based on scenarios that readily leap into our minds while we downplay risks of the options to which we are emotionally predisposed.

Tetlock (Tetlock & Gardner, 2015) poses a problem that highlights the tendency of even highly capable people to go with their gut reaction (System 1). I’ve used it with colleagues and to my surprise they did as I did – let “thinking fast” rule their response.

A bat and a ball together cost \$1.10. The bat costs a dollar more than the ball. How much does the ball cost? If you are like everybody else who has even read this famous question, you instantly had an answer: “Ten cents.” You didn’t think carefully to get that. You didn’t calculate anything. It just appeared. For that you can thank System 1. Quick and easy, no effort required. But is “ten cents” right?

The answer, of course, is no. Good decision making begins with working hard to avoid being trapped by System 1 thinking – and that’s hard because System 1 is always there and ready to provide the easy answer. The first characteristic of superforecasters is their constant effort to avoid letting System 1 take over. They do this in a number of ways, and Tetlock has enumerated them in the “Ten Commandments for Aspiring Superforecasters.”

Habits of Superforecasters: The Ten Commandments

Learning to be a superforecaster, or even endeavoring to improve forecasting capabilities is not an easy process, but it is possible, given time, effort and reasonable native talents. As a means of improving the performance of all of his forecasters, Tetlock (Tetlock & Gardner, 2015) developed a 60-minute presentation in which he enumerated and explained the key habits of the superforecasters on his team. This became known as the “Ten Commandments for Aspiring Superforecasters,” and according to Frank David of Forbes (David, 2015) should probably have a place of honor in most business meeting rooms. These Ten Commandments are listed and explained in the Appendix of *Forecasting* (Tetlock & Gardner, 2015) and introduced and discussed below.

- (1) **Triage.** We are all familiar with the process of separating the critical from the important from the nice-to-have. In making judgments, it is important to avoid focusing on the easy questions where simple rules of thumb can get close to the right answer or on the impenetrable questions (like where will technology be in 15 years), choosing rather to concentrate on what we believe we can ascertain about the proximate future. This implies gathering enough information to be reasonably confident about our assessments of the future. Tetlock cites two basic errors that can be made while attempting triage: failing to predict the potentially predictable or wasting time trying to predict the unpredictable. These two errors are like boundaries, and in doing triage the risks involved in the current situation should determine how close one is willing to get to either bound. Which error would be worse given the consequences for this situation?
- (2) **Fermi-ize – breaking seemingly intractable problems into tractable sub-problems.** We tend to be fairly good at this in our community, but perhaps we don’t use it often enough outside of constructing distributed simulations. The name derives from the playful habit of Enrico Fermi who, during his free time, would ballpark answers to puzzling questions like “How many extraterrestrial civilizations exist in the universe?” Tetlock uses the following as an illustration of Fermi-izing. A lonely gent in London wanted to guess how

many potential female partners existing in his vicinity: start with the population of London, then take the likely proportion of women (~50%) and the proportion of singles (~50%) and then by the right age range (~20%) who happened to be graduates of college (~26%) and of those the proportion he might find attractive (~5%) and finally the proportion of those likely to be compatible with him (~10%). He determined that there were about 26 likely candidates whom he could seek. All of the above uses rather crude probability estimates, but it was surprising how many of the Fermi-type estimates on the broken-down problem netted remarkably good estimates.

- (3) **Strike the right balance between inside and outside views.** This terminology can be a bit confusing, but the process is clearly defined. When considering a problem or situation, look for similar classes of problems and understand the salient parameters. Consider a problem we face routinely: cost and time estimation. You have a well-defined software project and you ask your team how long it will take them to complete the project. They begin by listing each piece of the project and the time and effort required to complete it (insider view). The time they estimate is most often optimistic, and frequently by a significant factor. Superforecasters would not spend time examining the details of the problem, but rather look for similar projects and from the time and effort expended on them, estimate a cost to complete for the current effort (the outsider view). The results will be more accurate than the detailed computation offered by the performers. On occasion, some team might be particularly effective and beat all odds, completing the project within the estimated time and cost. It is suggested that the individual team be taken as an exception and not included in the set of similar cases.
- (4) **Strike the right balance between under- and overreacting to evidence.** Tetlock calls this “belief updating.” The central issue is whether or not you integrate each new bit of data into your mental model of the situation and factors important to it. Updating assessments is important. “Savvy forecasters learn to ferret out telltale clues before the rest of us. The snoop for nonobvious lead indicators, about what would have to happen before X could, where X might be anything from an expansion of Arctic sea ice to a nuclear war in the Korean peninsula. Note the fine line here between picking up subtle clues before everyone else and getting suckered by misleading clues. Superforecasters are not super Bayesian updaters, but they are better than most of us. And that is largely because they value this skill and work hard at cultivating it.”
- (5) **Look for the clashing causal forces at work in each problem.** Respect counter-arguments. We each have our own beliefs, but it is important to actively accept that we might be wrong, even if only in this case. Consider a case of two different views. Bring them together creatively and form synthesis. It gets significantly harder where the number of differing views expands to a handful or perhaps a dozen. Synthesis is a fine art that reconciles irreducibly subjective judgments. It is difficult and without a rule-book, but it is critical in developing a nuanced view of the future.
- (6) **Strive to distinguish as many degrees of doubt as the problem permits, but no more.** The more degrees of uncertainty you can distinguish, the better forecaster you will become. Few things are either certain or impossible, and *maybe* is not very informative. In horse racing, sports events, medical procedures and many other fields, we are not satisfied with the notion that the outcome is 50-50. One would seek additional evidence to narrow the odds. Superforecasters take the time and effort to do this type of analysis routinely for any issue at hand.
- (7) **Strike the right balance between under- and over-confidence, between prudence and decisiveness.** Another type of risk assessment. Should you err on the side of jumping to conclusion to provide an immediate response or take what might be too much time in developing a solid, long-term assessment? Superforecasters understand the risks of both and routinely manage the trade-off between the need to take decisive stands and the need to qualify their stands. They must exercise their judgment enough to allow them to benefit, not just from the last error, but from the string of failures and successes, incrementally improving their accuracy.
- (8) **Look for the errors behind your mistakes, but beware of rearview-mirror hindsight biases.** Don’t brush your failures under a rug – own them and learn from them. Conduct unflinching post-mortems. This is the problem with lessons unlearned and not recording things that failed. We learn too little from our failures and tend to overlook flaws in basic assumptions. It is also critical to do thorough post-mortems on successes. Not all successes imply that the reasoning was right. Confidently reasoning along the same lines is a formula for future failure. Try looking for the documentation of the problems associated with JSIMS of JMASS or JWARS. How can we learn from our mistakes if we don’t have access to the honest evaluation of past problems?
- (9) **Bring out the best in others and let others bring out the best in you.** This is more than team dynamics; it’s team management with the end of producing the best possible product. In various parts of his text, Tetlock makes use of the habits of the Wehrmacht in developing war plans. Plans were laid out and everyone

in the unit, regardless of rank, was expected to criticize without repercussion from up the command chain. In many circumstances (some of which will be discussed in the following sections), bringing in a diverse group, understanding the arguments coming from different perspectives, being able to ask precise questions for clarification without alienating the other parties and being able to draw the fine line between providing helpful suggestions and micro-managerial meddling lead to the best judgments. This is certainly true in the process of developing requirements and engaging in the “never quite found” art of conceptual modeling.

(10) Master the error-balancing bicycle. The word *balance* is found in most of the commandments leading one to believe that good forecasting is a complex process of balancing opposing types of errors. This also means that the art has to be practice to be learned. Just as you can't learn to ride a bike by reading a physics book, you can't learn to be a superforecaster by reading a bunch of rules. They must be practiced with good feedback that leaves no ambiguity about whether you are succeeding or failing. This was the issue Kahneman brought before Tetlock. Nobody tracked the accuracy of the renowned pundits. Track results. Collect the data. Study both the successes and failures. And practice again and again, even if on simple questions at first. Like all other forms of expertise, superforecasting is the product of deep, deliberative practice.

(11) Eleven? Don't treat commandments as commandments. Use them as guidelines for developing and refining skills. Superforecasting requires constant mindfulness, even when dutifully trying to follow these commandments.

Do we all have to become superforecasters? No, but we do have to become better decision makers and we need to train our decision makers to be more creative and agile, able to balance risks of all types in the process of warfighting – whether that means major conflict, hybrid warfare, peace keeping, nation building or any of the many tasks we place before our military. We have to be better decision makers in choosing which technological dreams to attempt to realize in our systems. We have to be better decision makers as we work the requirements process, remembering that we are in the process betting on the future, just like the superforecasters. So, where do we go from here?

LESSONS UNLEARNED AND THE FUTURE

As I read Tetlock's work, several principles jumped out of the pages as being particularly applicable to our community. It seems inconceivable that before Tetlock's initial work (Tetlock, 2005), there had been no systematic study of the experts who repeatedly appeared in the media to make predictions about all sorts of things from political elections to the war in Syria and the presence of nuclear weapons in Iraq. They spoke with great confidence, backed by years of study and experience. Of course, other pundits commented on their perspicacity when they were right and put them to a certain amount of ridicule when they were wrong, but there was no consistent record keeping and lessons learned from the process. Once the data were collected, their track records were no better than random, than chimpanzees throwing darts. In considering past performance (both lessons learned and unlearned), it is wise to keep in mind two of Tetlock's commandments. Study past performance carefully to discover hidden as well as obvious flaws. When using the results, seek balance between avoiding errors and boldly striking forward into new technologies to address new capabilities.

Commandment (8). Postmortem your failures and your successes

Commandment (7). Strike the right balance between under- and over-confidence, between prudence and decisiveness.

Like the pundits on the media, we will produce capabilities some of which advance the art and science of simulation and training and some of which will fail spectacularly or expensively and we will repeat this process of unstudied performance until we do serious post-mortems on both successes and failures and make them available within our communities. Our lessons learned we study and incorporate into our toolkit, but the wealth of information from those failures is too often consigned to the archives or, worse, to the recycle bin only to have them reappear at some later time in another context to once again be a stumbling block.

The next sections are directed toward our modeling and simulation community as a whole, but, you contend, all the material I have presented is about individuals. True enough, but the practices translate. Micah Zenko, (Zenko, 2015) senior fellow at Chatham House and member of the McChrystal Group, “shows how organizations, not just individual, can overcome their biases toward false certainty and make good predictions, in geopolitics and business, in public and private sectors. With simulations, vulnerability probes, and alternative analyses that offer fresh eyes on a complex situation or intentionally oppose a certain position, red teams can greatly improve the accuracy of forecasts in the

same way that Tetlock's experts do." (Frick, 2015). The key, of course, is that management must commit significant resources to the red team and empower it to be brutally honest in its analyses. Tetlock's guidelines apply.

In his Fellows paper, Jack Thorpe (Thorpe, 2010) gifted our community in many ways as he recounted the roots of distributed simulation from the SIMNET project, but the part of the paper I found most appealing was the Appendix in which he sought the wisdom of his friends and colleagues. It is in these statements I found reference to those unlearned lessons; they will form the backbone of the sections on specific areas where we have failed to learn the lessons of the past.

The discussion of unlearned lessons is not intended to be comprehensive, but rather illustrative. They are in most cases limited to those areas in which I have been personally involved or at least closely acquainted, thus allowing me to provide evidence in which I have a reasonably solid degree of confidence. A major complicating factor in laying out the discussion is that while the technical areas are distinct, the issues that contribute to the fact that we have not really learned the lessons of the past are not. The discussion will begin with data and data sharing by focusing on the area of the physical environment – terrain databases and weather. These topics lead fairly naturally into the need for an appetite suppressant for increased resolution or granularity, most often sought under the umbrella of greater accuracy. The next technical issue in which there are numerous problems and lessons unlearned from other areas of modeling and simulation is the representation of human behavior and decision making, whether we are working within our own society or in the cultural milieu of other geopolitical regions. All of these technical problems lead eventually to the difficult process of developing requirements whether or not we think of the requirements as the official program REQUIREMENTS or simply requirements for a local simulation. In all of these areas, some of the major pitfalls could be avoided with coherent management processes informed by thoughtful conceptual modeling and incorporating one or more of the Ten Commandments.

DATA SHARING: HARDER THAN IT SHOULD BE?

Knowledge is power. Too often this translates into the idea that information is power and if I hold my information close, then I have a certain power – you are dependent upon me for the use of that information and ultimately you have to pay me for it. So, we make data proprietary or invoke intellectual property. But even when we cross those barriers and have a community interested and willing to share data, we collide with the problems of what the data looks like, how it is stored, and how it is interpreted.

The Senate calls out the importance of having a common data environment for modeling and simulation in its report on the National Defense Authorization Act (SASC, 2018) stating,

Modeling, training, and simulation efforts require significant amounts and different types of data in order to adequately simulate the operational environment. Each community and military service within the Department of Defense currently independently develops data for modeling and simulation purposes to support their own training, operations, analysis, test and development. The committee believes that this approach is not optimal, with independent data solutions increasing costs and inhibiting interoperability across the Department. The committee recognizes that some investment has been made to create capability to improve data sharing, reduce costs, and eliminate duplicative data collection and processing efforts. The committee further recognizes that additional investment and coordinated strategy may be needed to maintain and advance better use of data by the modeling and simulation community. The committee directs the Secretary of Defense to take actions to identify and address data collection, analysis, and sharing issues that are limiting development of more robust modeling and simulation capabilities.

While this statement recognizes the problems involved in sharing data across military services, there is also the issue of sharing data among contractors within each military service and among the Department of Defense and other agencies with missions requiring them to acquire and maintain data that are of use to the military, notably the National Geospatial Intelligence Agency (NGA) and the National Oceanographic and Atmospheric Administration. There are reasons why data is not easily shared among such groups. Some of the problems are embedded in law (Title 10, for the military), some in the funding and planning cycles for major acquisition programs, and in nearly all cases the lack of incentive to go the extra mile to share, because sharing involves solving technical problems. Discussion of the managerial issues is reserved for later in this section. For now, consider a few of the technical issues involved in sharing data.

DATA SHARING AND INTEROPERABILITY IN THE SYNTHETIC ENVIRONMENT

The Problems with Terrain Databases

The difference between a Map and a Simulation Environment is much like the difference between the novel *Phantom of the Opera* and the Broadway production *Phantom of the Opera*. Both contain the same “information” but one is a sensory, rich, and dynamically changing representation while the other is static and has little real value absent an intelligent reader (Numrich & Richbourg, 2016). What creates this richness is a layering of many diverse data sets, each with multiple formats and standards, many of which are developed, owned and maintained by different agencies.

Terrain databases begin with the creation of the elevation data for the targeted area. Elevation data itself are obtained from a variety of sources and then stitched together to provide what once were paper maps and today consist of a number of different digital formats in several resolutions. Creating a given digital format is not error free, and combining higher and lower resolution data in the process of developing a multi-resolution database adds to the problem. If a digital representation is to be turned into a paper map, many of these errors can be ignored; however, in today’s digital environment entities in models and simulations in addition to the plethora of applications dependent upon reasoning over the digital elevation data, small errors in the digital elevation data can cause significant errors in applications.

In the modeling and simulation world, creating the elevation data is just the beginning of building a terrain database. Figure 1(Numrich & Richbourg, 2016) illustrates layering data from different sources to build the final database. Whether the database is to be used for virtual or constructive simulations, the environment must contain road networks, and physical features like buildings, bridges, fences, vegetation and all other components that make the environment realistic to both humans and digital models. Humans are excellent at integrating across errors, but the models that operate in the digital domain must be perfect to be able to operate consistently.



Figure 2. Complex environment typical of today's simulation environments

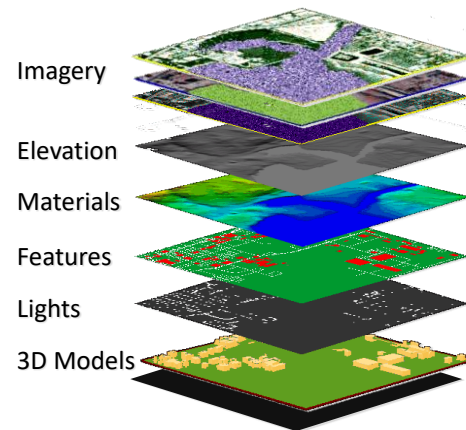


Figure 1. Layers combining to form a terrain database

Figure 2 (Numrich & Richbourg, 2016) is an image of the contents of a typical visual terrain database. The current emphasis on urban operations has increased the type and number of features at higher precision and resolution to support ever more complicated interactions in the simulation. Many of the issues involved in creating such representations for simulations were investigated in the process of developing the Synthetic Theater of War (STOW) program conducted nearly twenty years ago under the sponsorship of the Defense Advanced Research Projects Agency (DARPA) (Richbourg & Lukes, 2008). Yet today we still hear voiced the notion that “we are all using DTED data, so everything will work fine, right?” Lessons unlearned in the complexity involved in the construction and sharing of terrain.

We can use a couple of examples to illustrate the pathology that occurs when there are small errors in creating the simulation database out of the elevation data. Terrain databases may have used DTED (now an older format) data, but the actual database is most often developed using different geometrical shapes. If there is an error either in the original data or in their conversion into the new geoshapes, simulated entities may have difficulty reasoning across the digital data and behave oddly.

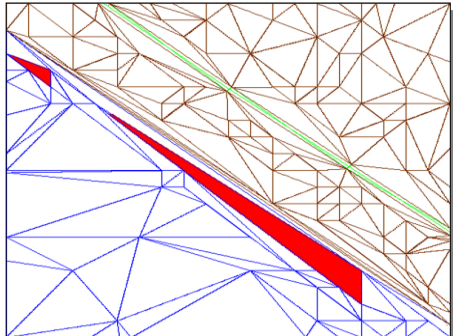


Figure 4. Red triangles indicate gaps in the terrain resulting from conversion of data format

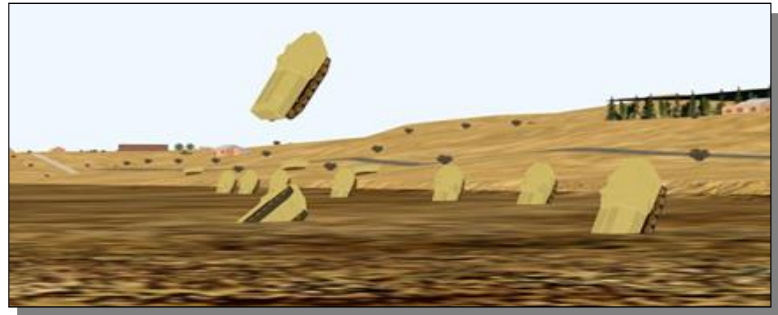


Figure 3. Strange behavior of tanks as they encounter the discontinuities in the terrain database

Figures 3 and 4 (Numrich & Richbourg, 2016) illustrate the pathological behavior of entities when errors occur in the simulation database. Generally one would prefer that tanks neither fly nor dive.

The above errors are technical problems in managing a single database. Errors compound when attempting to share databases across multiple simulations, all of which have been developed separately with their own individual databases. It is important to remember that the models recognize terrain features by the way they are defined and digitized. If we climb the layers from the elevation data through the materials and focus for a minute on the feature

Describing Content: Data Dictionary Lineage

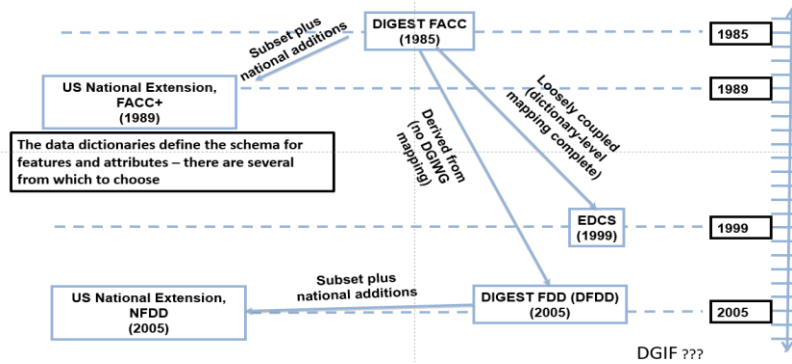


Figure 5. Pictorial history of major data models for physical features

database including such things as buildings, we encounter another set of problems. How do you represent a road? It is a highway, a residential street, a dirt road? What about a building? Certainly, it must have at least four walls, a floor and a roof, but how do I describe them? If we can get past the description of the feature itself, we still have the question of how it is pinned to the database, lest the White House be seen either floating in air or partially buried. Physical features added to the terrain are defined according to a data model which is described in a data dictionary. Unfortunately, there is not

just a single data dictionary; therefore, once again the simulation database developer is faced with the problem of translating. Unlike elevation data where the notion of “elevation” is reasonably well understood, physical models are defined in many ways using different terms in the definitions of the model. Thus, there must be semantic understanding or clear and unambiguous mapping between different data dictionaries. Figure 5 (Numrich & Richbourg, 2016) shows the lineage of several data dictionaries from the DoD and the NGA. Note that there are multiple endpoints and the “DGIF???” at the bottom right of the figure is the placeholder for the next generation. While all the products (the endpoints) can be traced to a common beginning, the fact that there are separate paths and endpoints is indicative of a lack of interest in building along the same lines to develop a single, continuously improving data dictionary.

At this point, it might be appropriate to ask if there have been success stories in sharing data, if not complete databases. Here is a positive answer with characteristics worth noting. The example is the Multinational Geospatial Co-

production Program (MGCP) run under the leadership of the NGA in partnership with countries all over the world to develop a repository of sharable elevation data. Each country had to adopt the mutually-agreed upon standards. The extent to which they could access the data in the repository was proportional to the data they delivered as long as their data met the standards and was deemed correct – relatively error-free. Each partner nation resources its own data collection and processing. There are a group of lead nations experienced in the geospatial production business that are willing to train and share capabilities with countries anxious to participate but lacking the training and experience with the data collection and processing tools. NGA has developed lessons learned after a number of years of working with the MGCP (Dellagnello, 2016):

- Search for an equitable quid pro quo
- Standardized production
 - Facilities fighting off the same map
- Relationships are critical
- Don't be bound by convention, history
- Patience is needed...keep a long-term focus...invest in the future

The key assumption is that no single partner has all the answers, and that by working together the partnership will realize change in the future that will profit all. Sacrificing individual progress to work cooperatively looks like Tetlock's Commandment (9).

Where Do Ocean and Atmosphere Fit?

The simple answer is that for the vast majority of simulations, they don't. According to Randy Garrett (Thorpe, 2010) "although environmental effects are a very significant factor for Intelligence (and Operations!), few systems incorporate weather time of day, or obscurants." Not only is the environment (ocean and atmosphere) difficult to model, once present, the models themselves have to be made capable of dealing with the effects. Each element of the physical environment included compounds the problems involved in correlations among the participating simulations in a multi-simulation or federated environment. Among the issues involved is the unfair fight:

- Unfair fights can occur when one (or more) simulation uniquely enjoys an "unnatural" advantage during interactions with other simulations (including "live" simulations)—such advantage can invalidate simulation results.
- Interaction between simulations using different synthetic representations of the same natural environment can result in unfair fights—there are several sources resulting from the sampling of the environment and use of the environmental factors by the models.
- Atmosphere and ocean are dynamic environments in which sampling of the environment and delivery of temporal changes must be consistent across all federates and all models in each federate sensitive to environmental effects to prevent degraded performance of a selection of models and systems.

Some environmentally sensitive models respond, not to the environmental conditions, but to environmental effects. For example, environmental modeling for aircraft using weapons from a significant altitude may rely only on a report of cloud cover, not on the relative humidity of the intervening air mass, nor any of the other factors that contribute to the formation of the cloud cover. For sensors operating in the undersea environment, knowledge of the temperature distribution with depth at various times of day and year are used to determine the manner in which sound, the primary sensor, travels through the water. In most cases, the sound transmission is calculated from environmental data and not provided as an environmental effect.

Thus, the use of air and ocean environmental data involves numerous decisions including:

- Does the simulation require environmental data or specification of environmental effects based on nominal environmental conditions?
- If environmental data is provided, to what degree are all the federated models sensitive to that data and is that sensitivity consistent across all federates?
- What degree of dynamism is required for the models in the simulation – leading to whether climatological or meteorological data is appropriate?
- Are there assumption in the models that constrain the type of data required?

Consider the following actual example of integrating the undersea environment into a simulation. In the quest for increased fidelity, the program manager sought water column information every six hours with the highest possible density of locations. The first question, based on the assumption that the manager was serious and understood the impact of the request was “which six hours” as the starting time of the samples makes a difference. Samples taken at 11:00 am may well have a very different profile than those taken at 3:00 pm if there is surface warming due to a bright sun. Asking that question launched a deeper discussion that eventually came to the question of what the models actually required. The model was based on a deep water approximation known as the reliable acoustic path. Neither the developers of the simulation nor the manager realized that the model was invalid outside a deep ocean context. The particular environment requested was for a shallow water basin where the model would have been invalid regardless of the data provided. This situation points to several problems:

- When the simulation was designed, the design process did not include experts in all relevant fields
- The data request was for the highest resolution, not the resolution required to address the problem
- There were assumptions in the models that constrained the validity of the output if the wrong environmental data were provided, i.e., high resolution environmental data

Data sharing is the first step, but the goal is to enable interoperability among simulations and the ability to recreate a synthetic environment to study environmental effects for systems under development and systems already fielded. The DoD definition of interoperability from the DMSCO glossary is “the ability of a model or simulation to provide services to and accept services from other models and simulations, and to use the services so exchanged to enable them to operate effectively together.” It is possible to operate effectively, but not produce valid results. An extension to this definition called meaningful interoperability was proposed in 2001 (Clark, 2001): “Additionally the models and simulations understand the context of the problem being addressed, the syntax and semantics of the data being provided and ensure the level of detail and fidelity is consistent across the entire federation without the need to develop custom hardware, software or tools (composability).”

The Battle Force Tactical Trainer (BFTT) program aimed at federating the individual training system aboard a ship to provide a realistic operational experience for pier-side exercises illustrates the need for meaningful interoperability. The individual trainers for the various shipboard systems were closed systems equipped with their own synthetic environments. Federating them entailed reconstructing them at the code level to allow all parts of the formerly closed system to interact with a common synthetic environment. After considerable furor on the part of the various contractors who developed the trainers, a common synthetic environment was provided by the maritime version of the STOW software.

Much of the work done to enable interoperability and data sharing in synthetic environments was done during the end of the 20th Century when DARPA was funding the STOW program and the Joint Forces Command resources multi-service experiments. As a result of the need to provide a consistent environment to shipboard systems, researchers in the Navy developed a server capable of receiving live data from Navy’s environmental data services and of using climatological data to design environments with features desired for specific training purposes. It linked effects capabilities, ensuring that all systems were using the data and effects simultaneously (Collins, 2005). The concept is shown in Figure 6. TEDServices imports the standard Navy data sources and builds the environmental data served to all federates by OASES. OASES can also deliver data to modules that compute effects needed by several of the federates, thereby guaranteeing that all systems of a specific kind operate the same way in the synthetic environment (ATLoS computes undersea transmission loss and IMPACT is an aerosol dispersion model). Today, very few developers know about environmental capabilities built during the STOW era and few integrate ocean and atmosphere into the models and federations. More lessons unlearned.

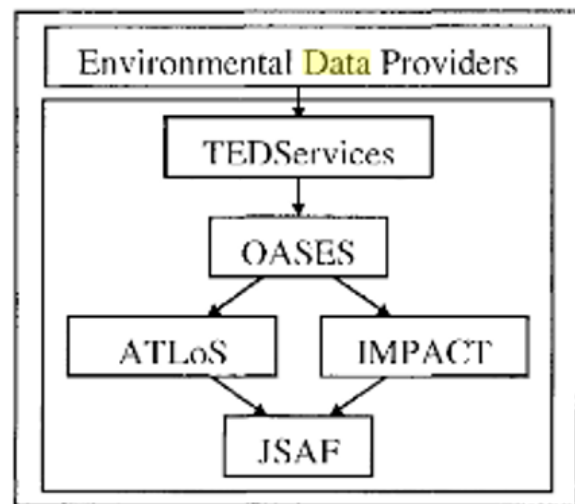


Figure 6. Environment data and effects server

DATA AND MODELING THE HUMAN ENVIRONMENT

Human behavior in a simulation has two primary components: the representation of the physical entity called a *human* capable of moving with or without complete articulation of body parts, and the representation of the decision-making process of the individual human or groups of humans. It is the latter that fascinates, challenges and frustrates modelers and simulationists.

The military simulation community recognizes SIMNET as the first giant leap into distributed interactive simulation. The SIMNET program started with a vision based on the founders' interpretation of the best blend of humans and synthetic representations. In writing about semi-automated forces and behavioral modeling, Thorp states (Thorp, 2010):

SIMNET was envisioned as a technology that allowed large numbers of humans to meet force-on-force and to compete driven by human intelligence and motivation. We did not believe that computer-implemented algorithms could mimic the internal workings of the brain, so the goal was to provide a playing field that connected humans and allowed them to do whatever they wanted to do, to include clumsy as well as inspired acts.

My initial introduction to human behavior modeling came in the 1990s in the form of the influence network modeling application called SIAM developed by Julie Rosen and colleagues. About the data available to provide estimates of human performance, Rosen states,

For situations where the measure of performance is a human's perception of the environments, little or no "hard data" exist. That is, the events and relationships that define a decision making process are burdened with uncertain and missing data. The key factors in addressing human decision making are diversity in the factors influencing relationships having both direct and indirect impacts; uncertainty in both the strength of the influencing impact between factors, as well as the likelihood that the factor exists; and dynamic nature of any list of significant factors and inter-relationships.

The preceding sections explored environmental data, data that can be extracted from the physical universe. Handling physical data in all its complexities is considerably easier than acquiring the information needed to characterize human performance. In attempting to set requirements for DIME/PMESII Modeling (Hillson, 2009), Hillson created an illustration of the types of data used in modeling and simulation. Figure 7 divides the data into four groups based largely on the nature of the data sources. Human decision making uses largely qualitative and transitional data, but in military simulation, those data have to be integrated and correlated with many other data sources, structured and unstructured. Thus, the issues of consistency and correlation present in the quantitative data for the environment are magnified and complicated when human decision making enters the modeling and simulation environment.

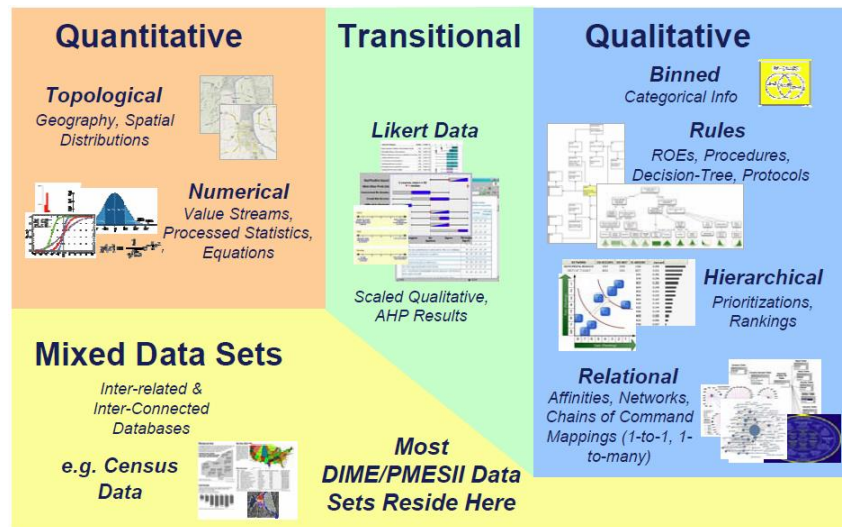


Figure 7. Data type used in modeling and simulation

While our capabilities in human performance modeling have made strides since 2000, the issues associated with the data and data sources remain. In fact, when we begin looking at smaller nations in areas where we are less familiar with the local populations and culture, data are far more difficult to find, interpret and trust. We attempt to remedy the paucity of data by moving to news feeds and social media for which we build data extraction and assimilation tools; however, the speed of ingestion does not compensate for the inferences the informed human can make from scanning articles of various sorts.

The lesson from SIMNET about the value of pitting human against human is continually being tacitly questioned as we build virtual environments populated by agents of various types reasoning according to different models in an effort to create realistic environments with responsive adversaries and allies. In his discussions with Thorpe, Craig Fields, former director of DARPA and member of the Defense Science Board, reflected on SIMNET and the modeling and simulation industry:

My biggest disappointment is that the pioneering work on large scale, many person gaming wherein people mimic people and software mimics physics, ala SIMNET, remains a footnote for DoD. The opportunity lost in terms of enhanced personnel performance; better decisions regarding what to buy and what specs and in what quantity; the evaluation of doctrine, rules of engagement, tactic and strategies in advance, and real mission rehearsal – what a loss!

One approach taken by the modeling and simulation community in the recent past is to focus on the gaming industry for games' ability to engage the participant and keep him (or her) coming back for more. It seemed to me that the training community was slow to adopt the notion of gaming for fear that there would be more than an acceptable level of negative learning. Games aren't in business to replicate reality to a given level of fidelity; rather they are aimed at suspending disbelief. So, the decision to use a game for training has to be assessed in the light of the skill being mastered. Distributed interactive games or, in the gaming world, the massively multi-player online games, provide added benefits of team formation and development in the presence of intelligent and adaptive adversaries. According to veteran board gamer, Peter Perla (Perla, 2017), "games are accurate precisely because they try to incorporate all of the external and human factors into the performance of a system. They can do this because they involve people..."

Wargaming includes everything from seminar wargames with little or no computer involvement to online games with rich synthetic environments. One of the most valuable facets of military wargaming is having "red team" players who take the role of the adversary after carefully studying his culture, potential interests and mindset. In the conclusion of his talk, Perla sets out the reasons why we use wargames.

We do it to help us all make more accurate predictions by leveraging all our combined knowledge, experience and creativity, so that we can make more effective decisions in complex and uncertain situations. We do it to question, to learn and to understand. We do it because Wargames entertain; they stir the imagination. Wargames engage; they stimulate the intellect. And Wargames enlighten; they create synthetic experience. And it is experience, both real and synthetic, that makes abstract risks tangible and effective planning possible

Before concluding that humans need to play humans, it is appropriate to return to why DoD wants to and needs to do some modeling of not only individuals but populations. Mike Van Lent (Thorpe, 2010) from SOAR Technology has an extensive background in games, and notes:

One of the principles of counterinsurgency is that the center of gravity of a COIN operation is the support of the local population... thus, effective simulation of COIN operations requires a realistic, culturally specific and reactive model of the population in the area of operation. Who does the population politically support? What are their needs and motivations? ...Modeling populations including the factions, tribes, and organizations within the population and the key individuals who lead or influence the population, is a critical next step to enable effective simulation-based training of the full spectrum of operations.

At this juncture it is important to remember Rosen's cautions about the availability of data to support modeling the desired human processes. Advances in extracting data from news and social media make it possible to unearth the reactions of segments of the population; however, understanding motivation and cultural drivers does not come primarily from these sources, but from individuals who study populations from the perspectives of anthropology, sociology, psychology and political theory. Without both information on reactions to current situations and in-depth understanding of the cultural drivers and internal social lines of fissure, the data paucity is too great to enable even the best models to provide reliable situation awareness

Thus, there are reasons why it is imperative that the modeling and simulation community continue to model human behavior and human decision making as long as we recognize that this, too, is an approximation to the decision making processes of a thinking, reacting and adaptive human person. Choosing when to use artificial intelligence (AI) and when to use humans is critically important if we are to rely on the results of our learning experience. Like Tetlock's superforecasters, we must weigh the risks in the process of opting for AI or human interaction. It is also important to engage in triage lest we spend too much time and resources trying to model the intricacies of decision making to a high degree of accuracy. This task should be consigned to the nearly impossible bin.

NEEDED: APPETITE SUPPRESSANT

Years ago, I remember being asked by a colleague to help present the concept of sharing environmental data across multiple programs as a means of providing high quality, common data sets to net an overall savings in time and resources while also enhancing interoperability. I was the guest and the presentation was being made to the most powerful Office Chief in Navy acquisition. We were in a small room with the principal at the head of a small conference table and the rest of his staff in chairs all around the room. I spoke a bit about the concept and he interrupted stating that he wasn't concerned so much about other offices and their needs, he wanted the highest resolution data available. At which point I rejoined, "Perhaps not." Silence enveloped the office and all of a sudden, he and I were the only two persons in the room and I was definitely on the carpet. People did not contradict him. But his approach was wrong and I felt compelled to point it out. I survived, and he had a somewhat more nuanced approach to resolution, fidelity and what his models and simulations might require.

Stating that "I want the highest resolution data," is an example of System 1 thinking. It's the easy answer, far easier than trying to discover what the right resolution is for the problem at hand. While it clearly costs more to acquire and deliver, the highest resolution data is not always the best for a given problem. Figure 8, taken from the tutorial on Fundamentals of Modeling and Simulation (IITSEC, 2017), shows the commonly accepted hierarchy of models from the most detailed at the bottom to the widest scope and least detailed at the top.

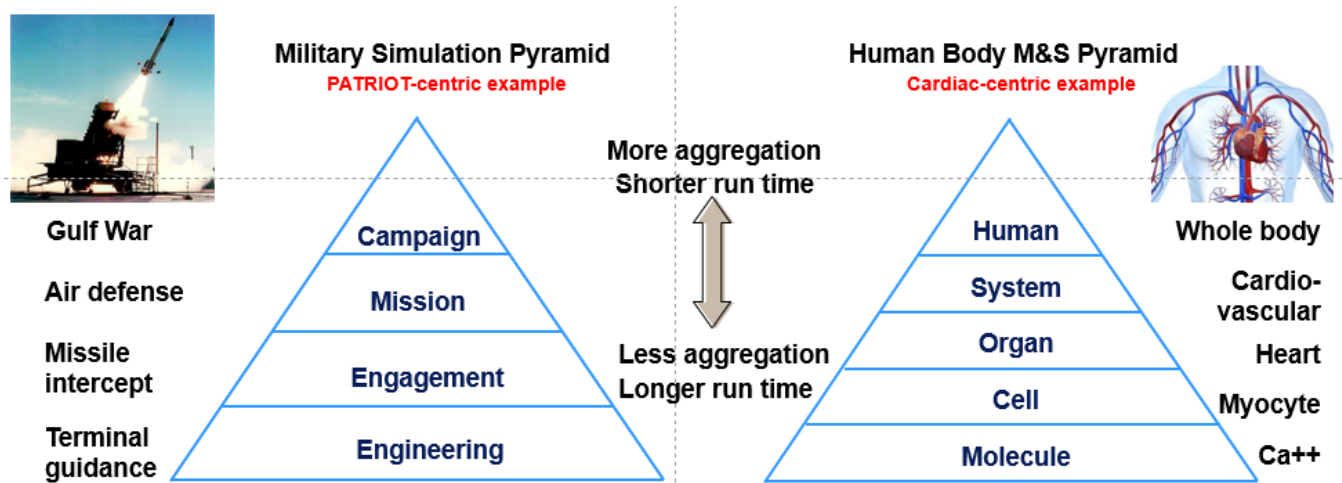


Figure 8. Hierarchy of models and simulations based on detail and aggregation

It does not make sense to attempt to model the cardiovascular system from the level of a molecule, nor is it reasonable to model the Gulf War from the engineering level of all weapons, platforms and systems involved. For example, moving from a 30-meter Digital Elevation Map (DEM) to a more highly resolved 3-meter DEM exacts a penalty of 100 points for each 30-meter marker. Furthermore, the cost in developing the highest resolution data, the size of the resultant database and the speed inherent in computing over the extent of data make very high resolution also very expensive in development and execution time.

The Joint Countermine Operational Simulation (JCOS) was a program developed to explore the utility of countermine systems in use and in development by Army, Navy and Marines when used in combination to provide access to the beach and beyond. The highest detail, therefore, was the area of operations, i.e., the near shore bathymetry and the beach where all the systems would be demonstrated.

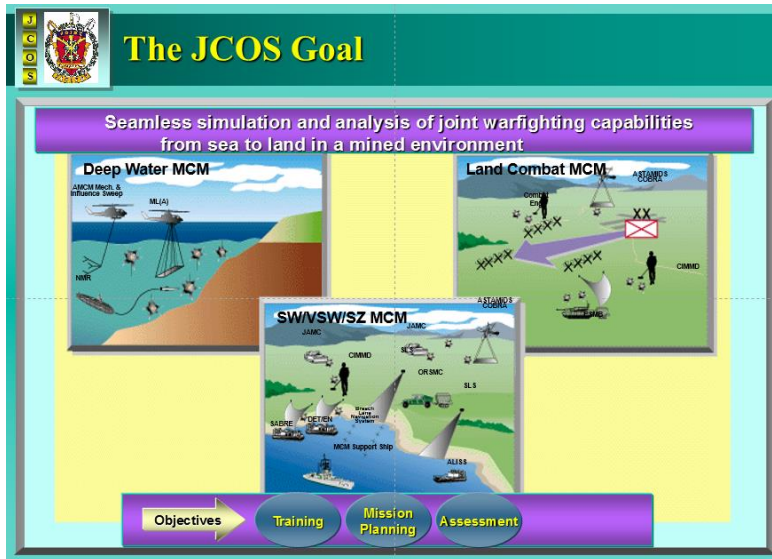


Figure 9. The environment of the Joint Countermine Operational Simulation

Figure 9 presents three views of the environment modeled in the simulation. Because a large number of systems were used in the shallow water approaching the shore, the JCOS simulation used the highest resolution bathymetry ever incorporated into a compact terrain database (database format use in the Semi-Automated Forces or SAF genre of simulations). In keeping with the resolution of the bathymetry and the shore where buried mines in minefields were represented, the forested regions on and behind the hills also contained high resolution representation of the physical environment – individual trees. As is typical for very high-resolution databases, it was finished just as the simulation was packed up to use for the NATO training exercise, MARCOT. We arrived in Newfoundland and prepared to run the simulation. All elements were loaded and

the start switch engaged. The simulation shuddered and stuttered for about 10 minutes before it gave up and crashed. After a considerable amount of time debugging, the engineers reached the conclusion that the database was too large to mount in the simulation due to all the high-resolution components. Database engineers started madly removing trees. There was no reason to have high detail in the forest – no systems were operating among the trees and vertical sticks were as valuable in the database as detailed trees. In fact, they were more valuable – with sticks the system loaded and the simulation could run apace with the actual exercise as was the original intent.

Bob Jacobs (Thorpe, 2010), an engineer from the SIMNET project, coined the term *selective fidelity*, but noted that “it cut against the operational community that had been lobbied by manufacturers to believe that only “full fidelity” simulation could deliver desired training results. But “full fidelity” itself is an illusive concept, as few could objectively define its attributes.”

Lessons learned in the use of appropriate fidelity in SIMNET and JCOS have not been learned. The initial reaction of nearly all users and program managers is to require the highest level of resolution available. High resolution is not synonymous with high fidelity. Low resolution can be “faithful” to the representation of the real world required by the model of simulation in question.

M&S MANAGEMENT, CONCEPTUAL MODELING, AND REQUIREMENTS

Many of the issues arising from lessons unlearned could be addressed by a more creative and consistent management of the Department’s modeling and simulation resources. Long-term planning is required to develop a consistent forward drive toward increasing capability, but while many studies point out the need for managerial leadership, the Department is loath to take up the challenge.

- Rudy Darkin (Thorpe, 2010). “Architectures should essentially be the government’s responsibility. Never outsource architecture. Architecture is the rules by which all the fast moving parts fit together. The trick is managing the architecture without stifling creativity – the same argument we often hear against standardization.”
- Jim Hollenbach (Thorpe, 2010). “Technical challenges remain (e.g., data engineering, modeling discipline, human behavior representation), but the biggest problem impeding the advancement of M&S are a dearth of DoD leadership and no ineffective DoD business Model.”
- Dr. Anita Jones (Thorpe, 2010). The DoD leadership should assert proponenty for interoperability of defense models and simulations. It is too important for the Department not to be proactive on this issue.

- Dan Kaufman (Thorpe, 2010). The government desperately needs to get away from proprietary systems. We need common libraries and standards to empower the end user.”
- Farid Mamaghani (Thorp, 2010). “While progress in this area [environmental data representation], and similarities in what is needed across communities, is notable on a technical level, there seems to be a lack of focused leadership and coherent strategy for tackling these problems across the board and under a uniform approach at the Defense level.”
- Robert Richbourg (Thorpe, 2010). “There are some challenges, too: for the MS&G architectures (e.g., HLA and DIS) there are not business of management models for DoD. There is a lack of leadership at the DoD level, and thus we have seen the emergence of different approaches without the needed interoperability.”

What is needed from the DoD leadership is not another set of REQUIREMENTS that hamstring and delay the execution of major MS&G programs, but a solid vision and articulation of enablers to achieve that vision. DoD leaders cannot achieve such a goal on their own – they lack the technical vision. MS&G program managers alone cannot provide DoD leadership with the vision – they lack the department-wide perspectives and interest because most of them reside in the Services. MS&G technologists are not the right community to provide the vision – their expertise and vision is in creating the new technology, not building and sustaining systems. Users can’t provide this vision because they are (in the words of Jim McDonough, USMC ret) (Thorpe, 2010) “captive of its [the operational community] experience and expertise: it will seek incremental improvements and miss the opportunity for breakthrough solutions.”

The approach requires groups of thoughtful individuals from various communities within and outside (notably the game development community) the Department. The individual would have to understand the background of DoD’s significant successes and failures and be able to bring creative postmortems of both as the foundation for learning from the past. They would have to be willing to humble themselves before the task at hand, listening to and understanding the diverse perspectives in the room. Their task is strategic, and thus, while learning lessons at levels of greater detail, they must be willing and able to articulate the clashing forces that prevent a single, coherent view and develop the appropriate paths through the pitfalls either attempting to solve the impenetrable problems or dwelling on the trivial. (Tetlock’s commandments 7, 5, 1, 8, 4, 9, and 3)

Impossible? Not really. It has happened either by *felix culpa* or deliberate planning when the Director of Defense Research and Engineering, Dr. Anita Jones, the SIMNET and follow-on programs at DARPA, and the Defense Modeling and Simulation Office were largely in sync and simultaneously supporting a vision of an interoperable suite of models, simulations and simulators. If this synergy was an accident, we can’t depend upon it happening again in the same way. Moving in this direction takes more than vision; it takes courage and purpose. But isn’t the security of this nation worth the effort?

That Sounds Like Building a Conceptual Model

Yes, and intentionally so. Whether or not we build conceptual models, we recognize that there is value in doing so. I recall being at a special meeting of MORS where there was a session devoted to gamers. These are the inveterate board gamers like Peter Perla who regularly develop board games for work or pleasure and as a community, they agreed that they would not begin to build a game until they had built a conceptual model. I was amazed considering how difficult it is to get members of our part of the MS&G community to develop even the most primitive conceptual models. Some of my colleagues who are not great enthusiasts about conceptual modeling regard it as simply part of good engineering practice.

At its core, a conceptual model is a concrete attempt to understand and describe a problem from the perspectives of all interested parties before casting that problem into code; as with the development of a corporate strategic direction, the conceptual model requires the thoughtful presence of all stakeholders. As in Figure 10, the conceptual model lives in two domains: the problem domain where the user can rely on the vision of the real world, and his experience with which to describe the problem that the simulation must address. The software engineers live in the model domain with an occasional nod to the problem domain. Project managers and software designers must live with a foot in each domain as they translate a system's desired attributes into the system

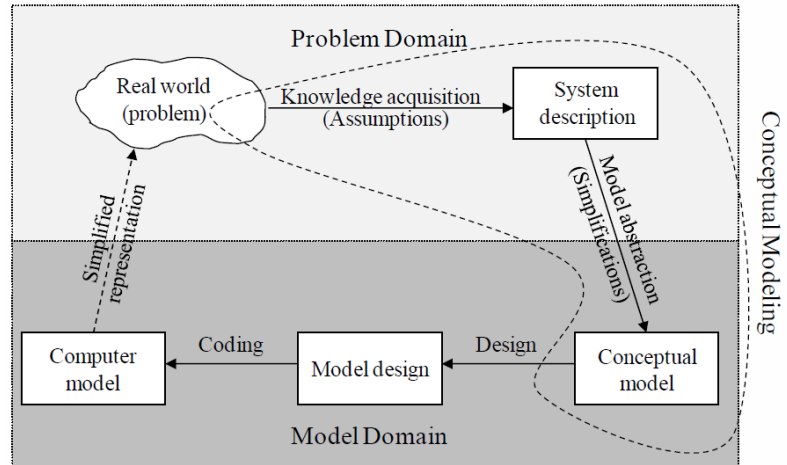


Figure 10. Robinson's image of a conceptual model

requirement—working all the time with the software designers to make sure that the desired behaviors can be implemented in the computer code. In general, the end user's role is to be firm about what the system has to do, while the designer dangles before him new and inviting technical capabilities. Meanwhile, the manager must keep in mind the resources available and add the voice of realism. Among them they must Fermi-ize, breaking down the problem into its components so that each piece can be examined under the bright light of triage, deciding which pieces are “must haves”, which are “not really needed” and which are “nice to haves.” Meanwhile, the technologist is pruning these lists from the perspective of the capabilities that are ready today, those in the mill and nearly done, and those like the holodeck – a gleam in the mind of the engineer. (Tetlock's commandments 1 and 2)

Meanwhile, the user will demand the product on a schedule that is the second cousin to non-obtainium. The software designer looks at the components and provides what will nearly always be highly optimistic (although not quite non-obtainium) and the project manager is seeking other similar projects to get a realistic idea of the time and resources needed for completion. The conceptual model is used to record all of these processes and most especially the compromises developed along the way to keep the project within the time and resources allotted. (Tetlock's commandments 3, 5 and 6)

The time has come to hand the conceptual model off to the software engineer who has been participating in the decision process with the model designer and is well aware of all the decisions and compromises made along the way. In the course of developing the concept model, all present interject lessons learned from prior work that has not gone well and efforts that have produced products with minimal issues and problems. Such lessons from the past are invaluable. (Tetlock 8) The documentation of the triage processes and the decisions made while developing the conceptual model constitute the initial steps needed for final validation.

Finally, the conceptual model provides the standard against which the product is considered acceptable. If, during development of the computer model, the software engineer discovers a new method that makes a desired capability not too hard, but viable, the conceptual model can be updated. It is a living document and part of the product throughout its lifetime; it is useful if a new program manager down the line seeks to an improvement that was non-obtainium. The software engineer can examine it, determine if there has been an improvement in the field that makes it possible or indicates that it is just as impossible as it was originally. This can save considerable time and resources otherwise expended in trying to create the impossible.

Wherein Lies Wisdom?

Strategic planning and conceptual modeling are smart ways to work, but they require resolve and courage in the face of expediency. The processes are closely linked to the commandments (actually best practices) from the Good Judgment Project. No, we don't all have to be superforecasters, but we do have the responsibility to exercise good judgment, and Tetlock provides guidelines for doing that. Do we as a community want to use these tools (strategic

planning and conceptual modeling with Tetlock's guidelines to improve them) to move our discipline ahead in the most effective way possible?

REQUIREMENTS or Requirements – Insurance or Impediment?

My background is in technology development and transition and I can't tell you how many times I have been slapped in the face with, "Where's the requirement for that?"

My stock response is that it just doesn't work that way. Basic and even applied research – the place where MS&G capabilities are born – are not done in response to a concrete set of requirements. In many cases, the basic research is most properly done to push forward the bounds of knowledge and if, in so doing, a useful product emerges, all the better. While research is generally about problem solving, the problems are not universally driven by requirements, and particularly not by REQUIREMENTS.

Requirements are the useful standards against which to measure performance stated as needed for a particular system. But there are also all sorts of other requirements. In fact, there is a whole group of intelligent officers and civilians in the Joint Staff whose role it is to maintain the requirement for acquisition programs. It seems that when it happens that someone should have done something and didn't, it becomes a REQUIREMENT and now everyone must do it.

There are some things in our field that should be REQUIREMENTS, and perhaps many other that are best left at the looser category of requirements. REQUIREMENTS are like laws, we make them and collect them and then rarely rescind them. Using standards should be a REQUIREMENT, but it remains a requirement if it is included at all in our contracting statements. Dan Kaufman would disagree declaring that we need common libraries and standards to empower the end user.

Rudy Darken (Thorpe, 2010) has already mentioned that architectures should be the job of the government. He adds that our contracting system and the way we work with contractors makes it hard to commoditize. By commoditize he means that we should encourage industry to add capability to an existing simulation tool (the commodity). Think about the explosion of the apps industry around the operating systems of our mobile devices.

It might be useful to apply Tetlock's commandments to the process of reviewing our REQUIREMENTS, deleting those that no longer make sense, and evolving an approach to add to them those that would add synergy in our community and products.

Lessons Learned: Coming Full Circle

While we don't have to predict the future, we do need to pay attention to the past, present and future in making our decisions. From the past, we could learn what to avoid; however, as a community, we don't catalog, red team and make these results easily accessible to our colleagues. We are a good bit better about our success stories, but we are not inclined to red team them to understand all the intricate facets of success. The more well-thought-out information at our disposal, the better we will do at planning—whether strategic planning for the community or individual program planning. It will help us avoid selecting a technology too soon or neglecting a valuable emerging technology. The more we share our knowledge, the more we will grow as a community. No, we don't need to become superforecasters, but by learning the techniques will we become a better and stronger simulation and training organization.

WHERE TO NEXT?

In their investigation of the computer's capabilities vs those of the human mind, the researchers at Deloitte (Guszcza & Maddirala, 2016) stated that there is both good news and bad news.

The bad news is that algorithmic forecasting has limits that machine learning-based AI methods cannot surpass; human judgment will not be automated away anytime soon. The good news is that the fields of psychology and collective intelligence are offering new methods for improving and de-biasing human judgment. Algorithms can augment human judgment but not replace it altogether; at the same time, training people to be better forecasters and pooling the judgments and fragments of partial information of smartly assembled teams of experts can yield still better accuracy.

One of the most innovative thinkers today is Thomas W. Malone, founding director of the MIT Center for Collective Intelligence. In his recent book, *Superminds: The Surprising Power of People and Computers Thinking Together*, Malone contends that groups of people working together, superminds, have been responsible for almost all human achievements in business, government, science and beyond. Malone would choose, not AI to replace thought, but hyperconnectivity, using networked computers to connect humans to one another at massive scales and in rich new ways. This could be a new mode of learning and training for our military leaders.

Today we often constrain our creativity by using machines to replace humans in parts of our business. We do excellent work adding robots to military operations to remove the human from the most intense danger. We build mechanical mules to carry heavy weights. We pilot unpiloted aerial vehicles from distant locations while collecting local intelligence in danger zones. We are getting more expert at using drones to work as teams to collect data in urban environments. But we rarely invite the machines to think with us.

Can you imagine a case in which machines with their rapid and copious memories link commanders in the fields and cultural experts to build understanding of the causal factors operating in a new and dangerous environment? What about linking 20 experts with the machines reading all the latest news and media reports in the area, providing translations in the process?

Rather than struggle to build responsive expert coaches, perhaps create a network of coaches, many coaches, who could work together with the learner. Machines can provide the accumulated information, and whatever Malone envisions as the rich environments (more than my mind can grasp now) with hundreds of human experts providing the nuanced judgments and intuition that we just don't know how to program.

Would it be heresy to say that perhaps we may have mined Roddenberry's communicator for all it's worth at this point, and that it's time to look to a new visionary like Malone for the future that is richer than the holodeck could ever be?

ACKNOWLEDGEMENTS

Anything I have managed to accomplish, I have done with the help of friends, often standing on the shoulders of giants not as yet recognized. They are too many to acknowledge individually, but I extend my gratitude to all of these, my colleagues. I must thank Dr. Randy Shumaker, Director of IST, who threw me into M&S as a member of the Navy M&S Management Team. I owe a debt of thanks to the military officers with whom I have had the honor to work: CAPT Paul Chatelier (USN, ret), CAPT Michael Lilienthal (USN, ret), CAPT Dennis McBride (USN, ret), CAPT Dylan Schmorrow (USN, ret), COL Michael Finnern (USAF, ret) and COL Kenneth Pieper (USA, ret). I owe a debt of gratitude for teaching me along the way to colleagues in the M&S community from DMSO, the environmental data community, from DARPA and JFCOM. I owe special thanks to NDIA and the IITSEC community for putting up with me for the past 2-plus decades, and most especially, to Ms. Barbara McDaniel who was the human face of IITSEC to so many of us. I wish to thank the Institute for Defense Analyses for continuing to support my involvement in this wonderful M&S community. And finally to those of you who had faith in me and felt I had something to say to the community from which I have learned so much. Thank you for this honor.

REFERENCES

- Brownell, C. (2015). If a retired birdwatcher can be a superforecaster, maybe you can too. Here's how. *National Post*, Don Mills, Ont. CA. Retrieved Sept 15 2018 from <https://business.financialpost.com/news/fp-street/if-a-retired-birdwatcher-can-be-a-superforecaster-maybe-you-can-too-heres-how>
- Clark, D.L., Numrich, S.K., Howard, R.J., Purser, G. (2001). Meaningful Interoperability and the Synthetic Natural Environment. In *Proceedings of the 2001 European Simulation Interoperability Conference*. Retrieved Sept 3, 2018 from www.sisostds.org/DesktopModules/Bring2mind/DMX/Download.aspx?Command=CoreDownload&EntryId=23049&PortalId=0&TabId=105
- Collins, J. & Scannell, C. (2005). Natural Environment Data Services in Distributed Modeling and Simulation. In Roy Ladner and Frederick E. Petry (Eds). *Net-Centric Approaches to Intelligence and National Security* (Chapter 8, p. 149), Springer US.
- Committee on Armed Services United States Senate (2018). National Defense Authorization Act of 2018. U.S. Government Printing Office, Washington, DC.
- Dellagnello, M. (2012). Multinational Geospatial Co-Production Program (MGCP). *JCIE Conference*, April.
- Dellagnello, M. (2016). Multinational Geospatial Co-Production Program (MGCP).
- David, F. (2015). Can Businesses Learn 'Superforecasting'? Easier Said Than Done, *Forbes.*, Oct. 4 2015. Retrieved Sept 17, 2018 from www.forbes.com/sites/frankdavid/2015/10/04/superforecasting/#193aac740785
- The Economist* (2015). Unclouded vision: Predicting the future. *The Economist*. Retrieved Sept 12, 2018 from www.economist.com/books-and-arts/2015/09/26/unclouded-vision
- Frick, D. (2015). Question Certainty. *Harvard Business Review*, Oct 2015, 130-131. Retrieved Sept 15, 2018 from <https://hbr.org/2015/10/question-certainty>.
- Galton, F. (1907). Vox Populi. *Nature*. Retrieved Sept 18, 2018 from <https://www.nature.com/articles/075450a0>. doi:10.1038/075450a0
- Guszcza, J. & Maddirala, N. (2016). Minds and Machines: The art of forecasting in the age of artificial intelligence. *Deloitte Review*, Issue 19, 9–25.
- Hillson, R., et. al. (2009). *Requirements for a Government Owned DIME/PMESII model Suite*. Washington, D.C., Office of the Secretary of Defense Modeling & Simulation Steering Committee. Retrieved Sept 17, 2018 from <http://drdeanhartley.com/HartleyConsulting/IWMetricOntology/n81%20DIME%20Report.pdf>
- Holcomb, W. (2016). How to predict the future. *PsycCRITIQUES*, 61(23).
- Hyndman, R. (2015). Exploring the boundaries of predictability: What can we forecast, and when should we give up? Retrieved from <http://robjhyndman.com/seminars/yahoo2015/>.
- IITSEC Fundamentals of M&S Tutorial Team (2017). Fundamentals of Modeling and Simulation. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (IITSEC) 2017*.
- Intelligence Advanced Research Projects Agency (2010). Aggregative Contingent Estimation (ACE) Broad Agency Announcement (BAA). Retrieved Sept 4, 2015 from www.airpa.gov/index.php/research-programs/ace/baa.
- Kahneman, D. (2011). *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.
- Malone, T.W. (2018). *Superminds: The Surprising Power of People and Computers Thinking Together*. Boston, MA, MIT press.
- Mazarr, M.J. (2016). Book Review: Superforecasting. *Joint Force Quarterly*, 81 (2nd Quarter, April 2016). Retrieved Sept 12, 2018 from <http://ndupress.ndu.edu/JFQ/Joint-Force-Quarterly-81/Article/702080/book-review-superforecasting>.
- Mellers, B., Stone, E., Murray, T., Tetlock, P.E. (2015). Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions. *SAGE Perspectives on Psychological Science*. 10(3), 267-281.
- Mlodinow, L. (2015). Mindware and Superforecasting. *NY Times Book Review*. Retrieved Sept 12, 2018 from www.nytimes.com/2015/10/18/books/review/mindware-and-superforecasting/html.
- Numrich, S.K. & Richbourg, R. (2016). Adding the “When” and “Where”: Basics of Environmental Representations. Tutorial. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (IITSEC) 2016*.
- Perla, P. (2017). Peter Perla on Prediction. *Keynote Address: Connections '17*. Retrieved Sept 3 2018 from <http://slideplayer.com/slide/14003660/>.
- Richbourg, R.F. & Lukes, G.E. (2008). Geospatial Data Quality for Analytical Command and Control Applications. *Proceedings of Critical Issues in C4I*. George Mason University, May 2008.
- Robinson, S., Arbez, G., Birta, L.G., Tolk, A., & Wagner, G. (2015). Conceptual Modeling: Definition, Purpose, and Benefits. In Yilmaz, L., Chan, W., Moon, I., Roeder, T., Macal, C. & Rossetti, M., Eds. *Proceedings of the*

- 2015 Winter Simulation Conference. Retrieved Oct 1, 2018 from <https://ieeexplore.ieee.org/document/7408386>.
- Rosen J.A. & Smith, W.L. (2000). "Influence Net" Modeling for Strategic Planning: A Structured Approach to Information Operations. *Phalanx*, 33(4). Retrieved Sept 15, 2018 from www.inet.leidos.com/docs/docs/inet-for-planning.pdf
- Surowiecki, J. (2005). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. Anchor Books.
- Tetlock, P.E. (2005). *Political Judgment: How Good Is It? How Can We Know?* Princeton, NJ: Princeton University Press.
- Tetlock, P.E. & Gardner, D. (2015). *Superforecasting: The Art and Science of Prediction*. New York City: Crown Publishing Group.
- Thorpe, J. (2010). Trends in Modeling, Simulation, & Gaming: Personal Observations About the Past 30 Years & Speculation About the Next 10. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (IITSEC) 2010*. Paper Number IF1001. Retrieved Sept 3, 2018 from www.iitsecdocs.com/search.
- Winter-Levy, S. & Trefethen, J. (2015). Pundits are regularly out-predicted by people you've never heard of: Here's how to change that. *Washington Post*, (Sept 30, 2015). Retrieved Sept 10, 2018 from www.washingtonpost.com/news/monkey-cate/wp/2015/09/30/pundits-are-regularly-outpredicted-by-people-youve-never-heard-of-here-how-to-change-that/?utm_term=.09a4b5009788
- Zenko, M. (2015). *Red Team: How to Succeed by thinking Like the Enemy*. Basic Books.