INSTITUTE FOR DEFENSE ANALYSES

Superforecasting: Proven Practices for Leveraging Human Ingenuity

S. K. Numrich

March 2020

IDA Document NS D-13127
Log: H 20-000102

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

About This Publication
This work was conducted by the Institute for Defense Analyses (IDA). The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Superforecasting: Proven Practices for Leveraging Human Ingenuity

S. K. Numrich, PhD, CMSP
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 AUTHOR

S. K. Numrich (Sue), PhD, CMSP, began her career at the engineering level of modeling and simulation and moved gradually into parallel and distributed simulation. She was selected by the Office of Naval Research, to represent the Science and Technology (S&T) community as part of the Navy's Modeling and Simulation Management Office. She developed and led a panel for The Technical Cooperation Program (US, UK, CA, AUS, NZ) in distributed simulation as an area of international interest and represented the US on the NATO Studies, Analysis and Simulation (SAS) panel as the simulation expert. 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 worked with the use of military simulation, the incorporation of human activity and behavior into simulations, and the validation of a variety of simulations. Sue joined the I/ITSEC community over twenty years ago and after serving on paper committees, she founded and was the first chair of the Tutorial Board. Sue authored four book chapters and over 50 technical papers and has had two academic appointments. A Fellow of the Acoustical Society of America, Sue was selected as the I/ITSEC 2018 Fellow. The research behind this paper was part of her fellow's presentation.

Superforecasting: Proven Practices for Leveraging Human Ingenuity

S. K. Numrich, PhD, CMSP
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.

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.

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 their 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. 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.

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.

Habits of Superforecasters: 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.

The following sections explore the application of the ten commandments of Superforecasters to some of the fields of modeling and simulation where mistakes of the past are continually rediscovered and lessons forgotten have to be relearned – lessons unlearned.

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 with which I have been closely acquainted. This allows 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 a number of the Ten Commandments. Look for the numbers in parentheses (C4, for example) that refer to the operative commandment(s).

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.

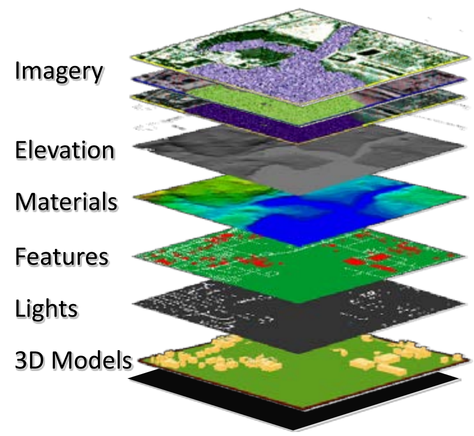

Figure 1. Layers combining to Form a Terrain Database

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. Layering is Fermi-izing (C3) applied to creating a terrain databases by breaking apart the problem into discrete layers.

Yet as recently as last year a developer of models intended for federation with or integration into other simulations dismissed the warning that there would be a problem with either process because of differences in the databases. The response to this warning was, "but everyone uses DTED, so there should be no problem."

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. Working to avoid discontinuities in data and models resembles the process of understanding diverse views and looking for clashes among those views (C5)

So, no, use of DTED does not guarantee that models will federate properly. The comment did come from a developer whose experience was with air assets and who was about to attempt integration with a simulation that represented ground forces and all the complications that entails in development of terrain databases. This outlook could be avoided by working from the start with a multi-service team (C9), guaranteeing that the richness of the terrain would not have been neglected.

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.

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? (C9)
- 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? (C6)
- What degree of dynamism is required for the models in the simulation – leading to whether climatological or meteorological data is appropriate? (C6, C2, C3, C9)
- 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. 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 (C9)
- The data request was for the highest resolution, not the resolution required to address the problem (C6)
- There were assumptions in the models that constrained the validity of the output regardless of the resolution (C1, C2, C9)

Another problem arises when live exercises depend upon climatological models. Live implies the here and now and in environmental terms, that means weather not climate. An environmentally tragic incident occurred when a live exercise use the best available climate models to make sure there would be no ill effect on aquatic creatures as a result of sound injected into the ocean. This was the right way to approach the initial design, but as luck would have it, the Gulf Stream which meanders was in a radically different position from the climatic average. Live sampling of the water temperature would have informed the exercise managers that something was wrong. Realizing that a live measurement was necessary to check that the model applied would have saved beaked whales from beaching themselves and dying. When using statistical results, it is truly important to check whether or not the average represents the current conditions. (C1, C2, C9) Lack of understanding could have been remedied by having a technical team assess live measurements on site during the exercise.

Figure 2 Environmental Issue Killed Whales and Created a Paperwork Burden for Future Exercises

As the military begins to model the elements of Multi-Domain Operations (MDO) with its emphasis on electromagnetics, cyber and space, the details of the atmosphere become increasingly important. The temperature gradient, differences in pressure, and the water content of the atmosphere have an effect on the propagation of electromagnetic waves in the atmosphere. Atmospheric effects alter the propagation path of the electromagnet waves and those changes affect the way sensors and receivers in military system operate. Computations in many simulation systems make the assumption that conditions in the atmosphere are uniform and consistent and that the propagation of electromagnetic waves obey simple optical propagation laws. Communication system, sensor systems, weapons and their targeting systems are all affected. Figure 3 taken from an online radar tutorial illustrates the response of electromagnetic waves to different environmental conditions.[1] These conditions can make a profound difference in where the waves propagate and the information that is returned along these paths.

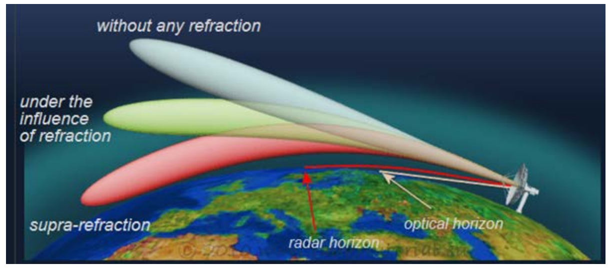

Figure 3 Illustration of Electromagnet Propagation under Differing Atmospheric Conditions

Once again, climate and weather play a role. Models using climatological data can be used to inform and instruct, but models using live weather must be used in exercises and certainly in live combat. Technical as well as operational understanding argue for the presence of a multi-disciplinary team in planning and execution MDO concepts. (C6, C9, C2)

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.

A lesson we have never quite learned is well-articulated by Julie Rosen, developer of the Situational Influence Assessment Model (SIAM). Rosen states (Rosen and Smith, 2000):

> "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."

One of the reasons our community continues to attempt to represent human cognition and decision making is the overwhelming presence of physical scientist in the simulation community. Too often engineers have devised programs to put some "science" into human behavior, neglecting the fact that physical laws and the methodologies consistent with them do not apply to the over-determined space of human thought and behavior. We expect numerical results with error bars where there is at best the indication of plausible outcomes. Because we fail to pay attention to the experts, we attempt to model exactly the imponderable. We also tend to be impatient with the social scientists who try sincerely to help us. (C1, C2, C5, C9)

NEEDED: APPETITE SUPPRESSANT

How often have you heard a user of program manager say, "I want the highest resolution you can get." This is often the results of "thinking fast," in cases where a study of the requirements has not been done in full, so get the highest resolution just in case. While it clearly costs more to acquire and deliver, the highest resolution data is not always the best for a given problem.

Case 1: Joint Countermine Operational Simulation (JCOS)

JCOS attempted to simulate countermine operations from deep water to the beach involving novel systems from the Navy, Army and Marine Corps. Because the littoral waters were important, the bathymetry developed in that simulation was the highest resolution ocean database in a semi-automated force simulation to date. To match the

[1] Publisher: Christian Wolff,Text is available under the GNU Free Documentation License, and the Creative Commons Attribution-Share Alike 3.0 Unported license, https://www.radartutorial.eu/07.waves/wa17.en.html

ocean, because it could be done, the hills behind the beach were modeled with a tree density representative of the shore of Newfoundland where the live exercise was to take place. As usual, the final simulation was put together in the last days before the exercise. At the exercise, every attempt to start the simulation failed – the terrain database was too large – too many trees. Trees were not needed for the exercise itself, but were added because…(C6)

Case 2: Marsupials with Stingers

And, yes, this did happen. As part of a helicopter training exercise, an attempt was made to provide more realism by including kangaroos in the operational area for the simulation. In the time-honored approach of reusing existing models, the kangaroos were modeled using infantry. The dispersed when the helicopters approached as required, but the pilots were astounded as the kangaroos reappeared over a ridge and launched a barrage of Stinger missiles at the helicopters. The exercise could have been completed without the added marsupials, but to enhance realism…(C6)

Case 3: JWARS

The JWARS program was an attempt to produce a consolidated campaign level model at higher resolution than the three (STORM, ITEM, and JICM) currently used by the analytical community. At every turn in the development, the community demanded greater resolution. JWARS did produce a functional product, but it is rarely used. The data required and the knowledge to produce it was more than the analytic community wanted to manage. Previous models had reached down to the service models to get the input parameters for the desired scenarios. Even when the data was provided, the model ran too slowly because it was computing too many of the results hitherto provided by subsidiary models.

In each of these cases, a carful exploration of requirements and computability would have pointed to an approach that did not require the highest level of resolution. Assembling a multi-disciplinary team with differing viewpoints could have sorted out the problems in time to avoid unnecessary expenditure of resources and unfavorable outcomes. (C1, C5, C6, C9)

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 (Thorp, 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 (Thorp, 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 proponency 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 (Modeling, simulation and gaming) architectures (e.g., HLA and DIS) there are no business or 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)

Role of Conceptual Modeling for Requirements

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. A conceptual model is a vision, a living document and part of the product throughout its lifetime; it is useful if a new program manager down the line seeks 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.

Development of a departmental vision for M&S would benefit from the employment of conceptual modeling at various levels: executive, program, and product. The membership of the conceptual modeling team would differ due to the nature of the problem space, but the approach would involve the decision makers prior to developing policy (executive level) rather than at the long and arduous process of coordination that often produces a product that is less than sub-optimal.

CONCLUSION

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?

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.

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?

REFERENCES

Brownell, Claire (2019) If a retired birdwatcher can be a superforecaster, maybe you can too. Here's how. *National Post*, Don Mills, Ont. CA. retrieved September 15 2018 from https://business.financialpost.com/news/fp-street/if-a-retired-birdwatcher-can-be-a-superforecaster-maybe-you-can-too-heres-how

Dellagnello, Marzio (2012) Multinational Geospatial Co-Production Program (MGCP), *JCIE Conference, April 2012*

Dellagnello, Marzio (2016) Multinational Geospatial Co-Production Program (MGCP)

David, Frank (2015), Can Businesses Learn 'Superforecasting'? Easier Said Than Done, *Forbes, Oct. 4, 2015, retrieved September 17, 2018 from https://www.forbes.com/sites/frankdavid/2015/10/04/superforecasting/#193aac740785*

Economist (2015) Unclouded vision: Predicting the future, retrieved September 12, 2018 from Https://www.economist.com/books-and-arts/2015/09/26/unclouded-vision

Guszcza, James and Maddirala, Nikhil (2016), Minds and Machines: The art of forecasting in the age of artificial intelligence, *Deloitte Review*, Issue 19, 9-25.

Hyndman, Rob (2015), Exploring the boundaries of predictability: What can we forecast, and when should we give up?, http://robjhyndman.com/seminars/yahoo2015/

I/ITSEC Fundamentals of M&S Tutorial Team (2017), Fundamentals of Modeling and Simulation, *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2017*

Intelligence Advanced Research Projects Agency (2010), Aggregative Contingent Estimation (ACE) Broad Agency Announcement (BAA), retrieved September 4 2015 from https://www.airpa.gov/index.php/research-programs/ace/baa

Kahneman, Daniel (2011), *Thinking Fast and Slow*, New York Farrar, Straus and Giroux.

Malone, Thomas W. (2018), *Superminds: the Surprising Power of People and Computers Thinking Together*, Boston, MA, MIT press

Mellers, Barbara, Stone, Eric, Murray, Terry, … Tetlock, Philip E. (2015), Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions, *SAGE Perspectives on Psychological Science*, Vol 10(3), 267-281.

Numrich, S. K. and Richbourg, Robert (2016), Adding the "When" and "Where": Basics of Environmental Representations, *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2016*, Tutorial

Rosen Julie, A. and Smith, Wayne L (2000), *Influence Net* Modeling for Strategic Planning: A Structured Approach to Information Operations, *Phalanx*, vol 33, No. 4, retrieved September 15, 2018 from http://www.inet.leidos.com/docs/_docs_/inet-for-planning.pdf

Tetlock, Philip E. & Gardner, Dan (2015), *Superforecasting: The Art and Science of Prediction*, NY, Crown Publishing Group

Thorpe, Jack (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 (I/ITSEC) 2010*, Paper Number IF1001, retrieved September 3, 2018 from http://www.iitsecdocs.com/search