MITRE | SOLVING PROBLEMS FOR A SAFER WORLD®

DECISION-SUPPORT TOOLS FOR NATIONAL POLICYMAKERS
FOOL'S GOLD OR TREASURE TROVE?

Christopher Ashley Ford, Ph.D.

 PRS 21-2877 1-11-2022

Executive Summary

Decision-support tools cannot replace the human judgment of high-level decision-makers, but they can assist national leaders in addressing complex national challenges. This paper offers a brief outline of one emergent decision-support methodology that is being developed and evaluated at the MITRE Corporation to support decision-making in advancing U.S. interests in our country's strategic competition with China. This MITRE methodology captures rigorously-structured insights from a community of human Subject Matter Experts (SMEs), and then incorporates these inputs into a Bayesian Belief Network (BBN) model. Beyond simply testing candidate Courses of Action (COAs) against this aggregated expertise, this methodology also permits users to adjust its parameters to test policy ideas against alternative environments. It thus has great promise in helping U.S. leaders cope with a challenging security environment.

Table of Contents

Figures

The Challenge

Decision-making is difficult for national-level leaders charged with solving problems in a global security environment of perhaps unprecedented complexity. On the best of days, policymaking routinely involves struggling with which decisions need to be made, frequently under tight time constraints, and then making them based on unavoidably ambiguous or contested information. Often, this causes leaders to employ various shortcuts or rules of thumb to help guide their behavior, sometimes relying upon unarticulated and unexamined assumptions about how the world works and the causal relationships between given policy interventions and subsequent policy outcomes.

ONE EFFORT TO HELP IS CURRENTLY BEING EXPLORED AT THE MITRE CORPORATION, WHICH IS WORKING TO IMPROVE DECISION-SUPPORT TOOLS THAT CAN INFORM MORE EFFECTIVE DECISION-MAKING IN FOREIGN AFFAIRS AND NATIONAL SECURITY PLANNING.

Under such circumstances, leaders often turn for advice to one or more Subject Matter Experts (SMEs), who provide input and advice informed by years of experience with, and study of, particular aspects of the policy problem set. Such SMEs can provide significant insights—at least if they are trusted by the leaders who employ them—but for many important high-level decisions, the sheer variety of expertise that may need to be brought to bear upon the question surely exceeds that which could be provided by any single advisor, however insightful and versatile, or even a small staff. No human is immune to error or bias, moreover, and both leaders and SMEs are rarely challenged to set forth and rigorously test the myriad assumptions that may lie behind their thinking, or to explore potential counterfactuals and their implications.

The American political system adds additional complexity to these dynamics, for we often populate senior leadership positions with persons who are not of the cadre of professional experts who follow the policy issues that these leaders will be asked to address. This can be a good thing, for such outsiders may bring "fresh eyes" to bear that are not hobbled by years spent marinating in the implicit assumptions and conceptual path-dependencies of received professional wisdom. But it can also be a problem, for "fresh eyes" are not invariably wiser ones, and where a leader lacks personal familiarity or experience with a particular problem and does not have access to enough SME input to make up the difference, poor decisions can follow.

So, are there ways we can improve the odds of making good decisions? Wouldn't it be better if leaders had access to a more expansive tool kit of decision aids? Should not such tools provide them with the ability to unpack SMEs' advice in detail, allowing leaders to assess its plausibility, identify its underlying assumptions, spot the areas of relative consensus or contestation encoded therein, test counterfactuals against received wisdoms, and explore the merits and demerits of alternative policy interventions in a rigorous and systematic way? More importantly—since to ask these initial questions is, I think, to answer them in the affirmative—is it actually possible to build such tools? Fortunately, the answer is "yes."

Please do not misunderstand. I am not suggesting we chase the pipe dream of discovering lawlike regularities in human interaction that would permit the development of a predictive, mechanistic "science" of human culture and facilitate forms of social engineering analogous to the crisp and quantifiable methods of mechanical engineering. I doubt any such thing is possible, and in any event, we are presently nowhere near anything like the mathematically-based "psychohistory" of the fictional genius Hari Seldon in Isaac Asimov's *Foundation* series. (As I recall those books, moreover, even that didn't quite work out as planned.) There is no escape from human judgment and subjectivity, and we would not be comfortable with one if it were offered.

Yet that does not mean we cannot aspire to exercise our human judgment better—in ways helpfully informed, but not hijacked, by more effective analytics based upon sounder and more plentiful data, and with both more rigor beforehand and more explainability afterwards. One effort to help is currently being explored at The MITRE Corporation, which is working to improve decision-support tools that can inform more effective decision-making in foreign affairs and national security planning. This effort is still in its early stages, and MITRE's tools are yet some way from maturity, but this methodology deserves attention as a glimpse of what is becoming possible.

The methodology described herein may seem complicated, but the basic concept really isn't. This is not an effort to cram the complexity and contingency of human interactions into an artificial mathematical construct that aims to replace recourse to traditional subject matter expertise and human judgment. Instead, it is a methodology for collecting expert human wisdom—and perhaps upscaling this collection through a sort of "SME crowdsourcing" that can incorporate input from a potentially much larger number of experts than one could feasibly fit around a conference table or manage at a meeting—while ensuring that SMEs are as clear and rigorous as possible in articulating their assumptions and reasoning. Thus, it uses and builds upon expert insights and human judgment rather than hubristically trying to replace them with a computer model.

Neither is there anything magical about this methodology, nor does it necessarily offer something a decision-maker could not, in principle, get from seeking input from a large number of SMEs in a careful and thoughtful way. Yet with the demands upon their time being what they are, no decision-maker at any senior level would conduct such a survey personally or be likely to structure the results and map subsequent decisions to policy advice in a clear, systematic, and explainable fashion. What this methodology can provide, however, is a means by which SME insights can be collected and used with maximum clarity and rigor, not only to aid specific decisions but also as an exploratory tool with which to test assumptions, perform sensitivity analyses of which factors have the most (or least) impact upon the situation, investigate the possible second-order effects of various alternative courses of action, and devise ways to maximize the chance of desired outcomes and minimize the odds of unwanted side-effects.

Building Models to Support Policy Decisions

The details of this methodology may seem a bit baroque, but it starts simply, with identifying a particular problem statement to be explored. (One should not underestimate this challenge, however. This initial step can take some time and effort, but it is essential to identify with care and thoroughness a discrete proposition to be explored.) Thereafter, it populates a series of software tools with relevant data and then employs these tools to help users (i.e., a policymaker) explore the anticipated merits and demerits of various candidate courses of action (COAs) in affecting that initial proposition.

For the sake of argument, let's assume that the starting proposition is, "The Kingdom of Nowehr follows China's lead at the United Nations." (Such a proposition could be true or false by degrees, but it's the statement we wish to evaluate and the policy outcome we wish to affect, either positively or, presumably here, negatively.) The model-builders might then identify a series of candidate causal relationships, such as "more Chinese loans to Nowehr make it more likely to follow China's lead" or "giving the Raja of Nowehr a state visit to Washington, D.C., makes that less likely." Additional candidate assumptions may also be set forth about other causal relationships, such as "Nowehr's persecution of dissidents makes a state visit to Washington less likely," "the Raja's speech on climate change at Davos makes USAID assistance for Nowehr more likely," or "expansion of Nowehr's GDP makes the Raja more likely to seek proposals for an expanded intermodal transportation hub at the Port of Nowehria."

Altogether, there are likely to be a great many of these candidate assumptions, which are loaded into a software tool (in this case, an internal MITRE application called "Loopy") that constructs a causal map of the relationships they indicate. Each proposition is a "node" in this map, and each causal relationship is represented by directional arrows running from each node that has an effect upon other nodes. **Figure 1** illustrates a simplified version of this example's Loopy map. In actual use, the process will likely create a large map consisting of different nodes (circles), connected by a web of causal relationships or "edges" (arrows).

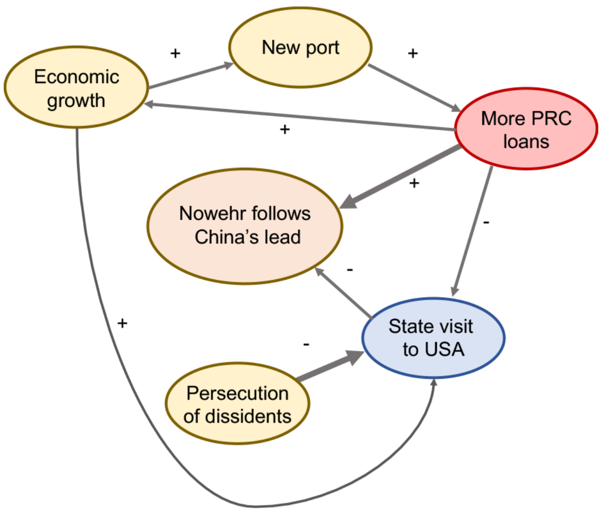

FIGURE 1. SIMPLIFIED LOOPY MAP EXAMPLE

These connections are initial working assumptions that need to be evaluated. Accordingly, the modeling and analysis (M&A) team consults SMEs who provide input to help vet these candidate nodes and edges (i.e., propositions and relationships). Perhaps the SMEs will reject one of the proposed connections, such as shooting down the previously mentioned idea that a speech at Davos is likely to be at all relevant to USAID decisions. Or perhaps the experts will add new ideas, such as that developing a new port infrastructure at Nowehria is likely to result in the Raja seeking more Chinese loans, or that economic growth is likely to make the prospects of a state visit to Washington, D.C., more likely. Additional SME insights are gathered as experts "play out" out these connections through role-playing exercises as stakeholder entities in the scenario (e.g., in this case, such as the Nowehria Port Authority, the Sultan, President Biden, the U.S. Department of Defense, the Development Finance Corporation, or the Bank of China). This is done using another MITRE tool known as "Serious Game," through which they interact with each other in exploring what they feel to be the parameters of possible node/edge relationships.[1] This role playing exercise brings about adjustments to and revisions of the Loopy map, likely increasing its fidelity and subtlety as a result of expert input.

Note that the arrows generally indicate a causal connection and its directionality (e.g., that persecuting dissidents affects the chances of a state visit) but that the weighting of the arrow-lines can be used to signal connections that SMEs have suggested are likely, in rough terms, to be especially powerful. Note also that this figure now incorporates "+" and "-" symbology to denote whether a given causal connection has a positive or negative effect. Read together, for instance, the combination of arrow weights and positive/negative notations signals that the economic dependency created by Chinese loans is likely to make Nowehr particularly likely to follow Beijing's lead at the U.N., that persecution of Nowehri dissidents is particularly likely to preclude a state visit to Washington, D.C., and that such a visit (if it were to occur) is likely to make Nowehr only somewhat less likely to follow China's lead.

The next step takes the depth and sophistication of this unfolding cognitive map to a further level by doing a deeper expert-informed tuning of these node/edge connections. Using an internal MITRE tool called Descriptive to Executable Simulation Modeling (DESIM), members of a group of SMEs (of essentially any size) each answer a lengthy series of pairwise questions about the relative importance of various factors or events. In each question pair, an expert is asked to specify the degree to which a given connection is felt likely to be true. One question might be something like, "All other things being equal, how likely is it that Nowehr's effort to construct a new port infrastructure will usher in new Chinese loans?" Experts answer two types of questions: (1) whether a particular proposition is valid; and (2) what the relative strength or importance is of each pair of valid propositions. A slider is used to indicate these relative strengths. This is not meant to be a rigidly precise assessment, but merely to identify and approximate relative degrees of perceived certainty. Responses are collected from each expert for each of the (potentially many) arrow connections represented in the expanded and adjusted Loopy map.

Note that this step is not asking the SMEs to recommend any specific overall course of action, a decision in which all manner of factors might come into play—as well as higher-level value judgments—not all of which may be within the ambit of a given expert's particular expertise. Instead, it restricts its queries to highly granular questions about those areas in which a SME is likely most authoritative and reliable: what factors are likely to affect what factual outcomes, in what ways, and to what degree.

[1]At present, this gaming is done through a web-based interface that allows asynchronous participation between six to ten participants. All decisions within the game are stored in a database for potential future analysis and integration into other analytical tools.

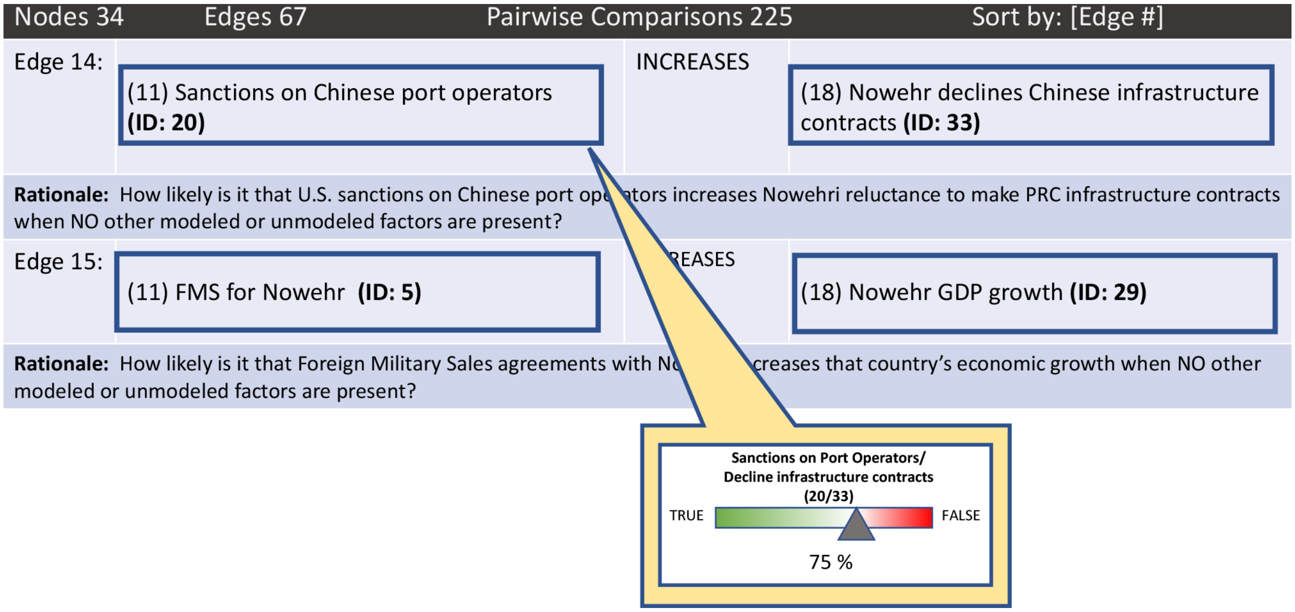

FIGURE 2. SIMPLIFIED DESIM CAUSAL WEIGHTINGS EXAMPLE

Figure 2 illustrates a simplified version of this example’s DESIM output. This particular image is largely populated with nonsense data, but hopefully it still serves to illustrate the sort of pairwise questioning presented to SMEs during this process, as well as indicating the way their responses are captured in the form of general percentage weightings. (The image suggests what an expert might see when in the process of sliding the grey arrow toggle to indicate his or her assessment of how likely it is that sanctions on Chinese port operators will depress Nowehri interest in getting port infrastructure contracts from China. In this case, the hypothesized expert feels this proposition to be highly likely: “75 percent.”) Each SME will be asked large numbers of such questions, and as many experts can be consulted as the model-builders want: the software allows near-infinite scalability.

The DESIM software collects and aggregates all SME input, creating a network of directional relationship strengths for each arrow in the Loopy map that reflects input from all surveyed experts. In the next step, this information is transferred—currently by hand, but hopefully in the future more automatically—into a commercially-available software tool called “QGeNIe”[2] to produce a version of the cognitive map in which are embedded all these the DESIM-derived causal weightings (expressed in percentages).

But this is not simply a fancy way of displaying the collective conclusions of a group of SMEs. What QGeNIe produces is a Bayesian Belief Network (BBN)—a probabilistic graphical model of what is known or presumed about a specified set of relationships—with easy interfaces that allow a user to walk through the various assessed causal connections therein to evaluate a given COA’s expected probability of affecting the starting problem

[2]QGeNIe is an interactive model building, learning, and exploration tool with an intuitive graphical interface. It is the flagship product of BayesFusion, LLC, an artificial intelligence modeling and machine learning company. See https://www.bayesfusion.com (visited October 25, 2021).

proposition (e.g., Nowehr's likelihood of voting as Beijing says at the United Nations, given that one or more of the propositions associated with a causal node is judged to be true).[3]

In the fictitious example, there are only a handful of nodes and relatively few connections; conclusions on this scale are likely still to be clearly accessible to (and articulable by) the "naked mind." Yet this same methodology can be employed to create a more complex BBN having many nodes and even more connective arrows, with each node linked to many others with varying conditional probabilities. Alternative "pathways" from any given node through the maze of nodes and edges to the foundational problem question can thereby be mapped and assessed on the basis of the probability outcomes assigned—on an aggregated basis—by DESIM's inputs from the SME community. This allows users to evaluate different policies, or packages of policies, on the basis of their chances of weaning Nowehr from the role of China's U.N. lapdog.

Moreover, as implemented in QGeNIe, the BBN provides a prescriptive capability. It facilitates the evaluation of potential courses of action by allowing the user interactively to alter the probability weights assigned to any of the cause-and-effect relationships set forth in the model, so as to influence the likelihood that the focus proposition representing the foundational problem (here, the proposition that the Nowehris follow Beijing's lead at the United Nations) will be true. In effect, this permits the user to discover the impact upon a focus node of policy decisions that

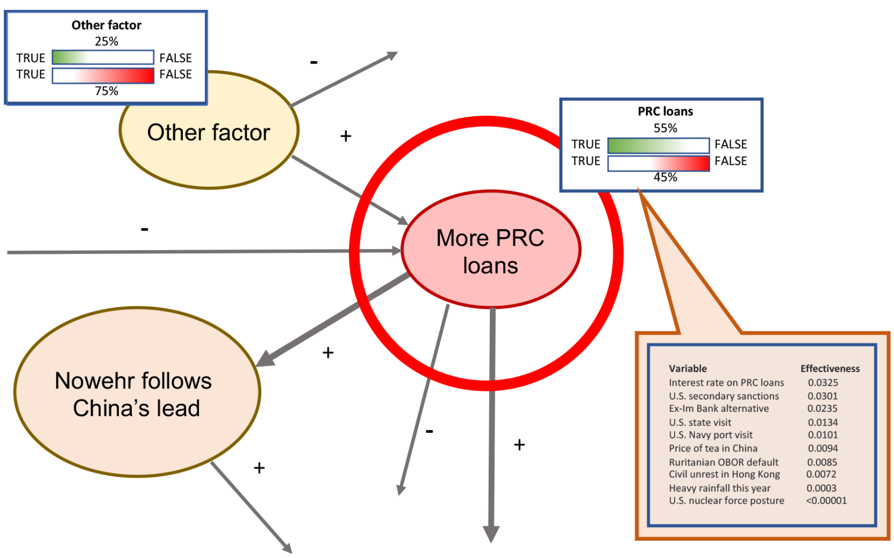

FIGURE 3. SIMPLIFIED BBN MODEL EXAMPLE

[3]A Bayesian Belief Network is a Probabilistic Graphical Model that represents conditional dependencies between random variables through a directed Acyclic Graph in order to understand the structure of causality relationships. (Conditional probability is the probability of a random variable when some other random variable is given.) See, e.g., Atakan Güney, "Introduction to Bayesian Belief Networks," *TowardsDataScience* website (November 20, 2019), available at https://towardsdatascience.com/introduction-to-bayesian-belief-networks-c012e3f59f1b. Such graphical models help discover and describe causality rather than merely identifying associations, as is the case in standard statistics and database technology, and can thus be useful in predicting the probabilistic effect of interventions. See, e.g., Sally I. McClean, "Data Mining and Knowledge Discovery," *Encyclopedia of Physical Science and Technology* (3d ed.) (Cambridge, Massachusetts: Academic Press, 2003), at 229-46.

make causal node propositions more (or less) likely to be true. (You can, in other words, test what would likely happen with different policy inputs under different sets of causal assumptions.) QGeNIe provides guidance as to which nodes are assessed likely to have the greatest effect on the focus node, with a list of other nodes ranked by their impact in a separate window. **Figure 3** illustrates a simplified graphical form of the BBN model.

This image is populated with still more invented data, but it should, at least, convey that with a QGeNIe BBN populated with DESIM data, the user can click any given node and see what the collective input of the SME community thinks about the relative chances that alterations will affect outcomes. This display indicates, for instance, that while additional Chinese loans are more likely than not to increase Nowehr's propensity to follow Beijing's lead at the United Nations, American use of secondary sanctions as a pressure tactic is likely to be more effective in affecting the extension of such loans than would be the alternative choice of making changes in U.S. nuclear force posture, but that neither of these policies is likely to be strongly effective.

Conclusion

To those accustomed to computer models offering the predictability and clarity of the solutions commonly seen in traditional mechanical engineering, the subjectivity inherent in building this decision-support tool might seem unsettling. Yet I suspect that this subjectivity is actually a great strength.

On one level, using this methodology is no different from—and should be no more unsettling for any policymaker than—the routine step of asking a seasoned SME about the implications of an idea. Yet it takes things several important steps further, not only by capturing that expert's best assessment of the situation but also by (1) forcing assumptions to be articulated as explicitly and clearly, and with as much granularity, as possible and (2) providing a mechanism through which insights from a very large number of additional SMEs can be captured and made simultaneously available to inform decisions.

Because users can explicitly identify and (if they wish) adjust the SME assumptions encoded in the model to evaluate how specific changes might affect outcomes, a range of hypothesized models can be compared. Moreover, if there is curiosity or disagreement about whether a particular causal connection exists, about its directionality, or about its strength (i.e., the probability weight given it), the model(s) provide a useful platform for articulating differences of opinion and perhaps reaching a consensus. Such "tunability" helps to provide at least a partial response to the classic "N of 1" problem of the social sciences: the fact that in understanding and seeking to draw lessons from large-scale, real-world social phenomena, we usually only have a single "run" of data to learn from. (You cannot rerun history to test your hypothesis.) This methodology doesn't really allow for true experimental repeatability of the sort so common in disciplines such as physics and chemistry, of course, but it does at least permit experimental counterfactual analytical "probes," as it were, that may cast light upon how alterations in underlying assumptions may produce alternative outcomes.

More straightforwardly, another advantage of this methodology is the "explainability" of results that it can offer as a result of the expository rigor of forcing participating SMEs to focus upon the particular pairwise questions that are posed and to provide a "best guess" of the specific relative weight for each response. The method looks complicated and can be time-consuming, but it permits every element, assumption, and linkage to be explicitly identified—and for each to be explained, or perhaps second-guessed, if questioned.

Naturally, as with the advice given to a decision-making national leader by a staff advisor, this methodology only offers advice and suggestions. Leaders can accept or reject such advice using their own judgment. But the explainability that is baked into this methodology ought to add to the "trust factor" associated with using it, making decision-makers more confident using such tools in framing and exploring the likely implications of alternative policy choices.

To be sure, this methodology, as described here and based upon the current "work in progress" status of its development at MITRE, still has some drawbacks. Perhaps most obvious is the laboriousness of the preparatory work that presently goes into it, particularly in the form of the hand-crafted model-coding that is still needed to pull it together and the many hours of SME time it takes to prepare and populate a suitably large collection of pairwise DESIM responses. Nonetheless, this methodology still has great promise, particularly for decision-makers (e.g., policy and strategic planning staffs) with the ability to focus upon broader, overall policy directions without being overwhelmed by the "tyranny of the urgent" that so often understandably preoccupies operational offices.

There is no magic solution here, and no leader should simply outsource critical judgment to an algorithm or model, however clever a construction it might be. (Even when good models exist, it can be valuable to study a problem—where time permits—using multiple methods to explore whether and where they converge or diverge. Nor in such a context should users trust a model's output reflexively; understanding how a model works is critical to appreciating when, and to what degree, it is likely to provide valuable insight.) Nevertheless, this BBN-based methodology may be very helpful in supporting sound decision-making, by forcing more clarity and rigor upon SME expertise, by assisting in the identification of higher-order effects or other non-obvious implications of SME "mental models," and by making chains of causal reasoning explicit for, more accountable to, and more (in a sense) "testable" by the leaders who rely upon them in trying to meet the security challenges of a threatening global environment.

ABOUT THE AUTHOR

The Hon. Christopher Ford is Distinguished Policy Advisor at MITRE Labs and a Visiting Fellow at Stanford University's Hoover Institution. Dr. Ford previously served as U.S. Assistant Secretary of State for International Security and Nonproliferation, also performing the duties of the Under Secretary for Arms Control and International Security.