

APPLYING METHOD

TO MADNESS:

A USER'S GUIDE

TO CAUSAL INFERENCE

IN POLICY ANALYSIS

Jessica D. Blankshain

Andrew L. Stigler

Jessica Blankshain and Andrew Stigler attempt to make the analytical tools frequently used in social science research more "user friendly" by explaining what it means to investigate causality. By providing a reader's guide to social science and policy analysis, they hope to enable practitioners to make stronger contributions at all levels of policymaking.

When national security practitioners — military and civilian alike — encounter academic social science, often in the context of professional military education, they usually respond in one of two ways. The first is with deep skepticism, sometimes bordering on antagonism: All this academic theory is nice, but it doesn't have anything to do with the real world where I do my job! Why should I believe this analysis? You can make numbers say anything you want! The second is with uncritical acceptance: I don't understand the math, but it was published by people with Ph.D.s so it must be true! This theory predicts X, so that's what will happen, right?

On the other side of the classroom, faculty like ourselves similarly struggle to communicate social science findings and approaches to students who may or may not have the academic training to critically evaluate them. Should we simply not assign any empirical social science or tell our students to "skip the math" and just read the introduction and conclusion? Or should we assume they have graduate-level training in social science methods and proceed with our lessons, ignoring the blank looks that greet us? Or should we do our best to turn our students into full-fledged academics in training?

Other professional military education dynamics exacerbate the challenge of teaching social science and other sophisticated analytical work. The students in any one classroom often have a wide range of experience, both educational and operational. One student might have a Ph.D. in a social science field, while another has an undergraduate degree in engineering and hasn't been in an academic classroom in decades, and yet another has a professional degree in law or medicine. Making things even more difficult, many programs operate on a compressed time frame, for example, conducting a full master's degree program in 10

months. More academic exercises in policy analysis and critical thinking compete for space with the need to cover ever-changing information on emerging threats and operational realities.

In addition to structural constraints, there are cultural challenges.¹ Many military students seek clear, definitive answers to questions, while academics are used to raising more questions than we're able to answer. Students are often uncomfortable with the idea that social science is still science even if it often can't produce exact predictions. And as Tami Biddle has noted, military practitioners "tend to be skeptical that theories produced by academics can help them understand war, which they believe is their dominion. After all, academics dwell in the realm of the abstract and the theoretical while military professionals dwell in the realm of the concrete and the real."² While this tendency may be particularly acute among military professionals, civilian practitioners are not immune to it. But simply ignoring academic social science — including social science methods — is not a productive solution. Social scientists use a range of tools to investigate hypothesized causal relationships *in the real world* and create generalizable knowledge that can be applied to other situations. Familiarity with sophisticated causal analysis is a key part of preparing practitioner students for careers in national security decision-making positions.

In their 2020 vision for professional military education, the Joint Chiefs of Staff argue that military success

cannot be achieved without substantially enhancing the cognitive capacities of joint warfighters to conceive, design, and implement strategies and campaigns to integrate our capabilities globally, defeat competitors in contests we have not yet imagined, and

1 For more on this, see, Joan Johnson-Freese, *Educating America's Military* (Abingdon, UK: Routledge, 2013).

2 Tami Davis Biddle, "Coercion Theory: A Basic Introduction for Practitioners," *Texas National Security Review* 3, no. 2 (Spring 2020): 94–109, <http://dx.doi.org/10.26153/tsw/8864>.



respond to activity short of armed conflict in domains already being contested.³

They call for an increased focus on “[c]ritical strategic thinking,”⁴ which would seem to include the ability to critically evaluate information and make educated decisions about the likely consequences of different actions or approaches. Students cannot achieve the Joint Chiefs’ requisite joint professional military education goals, such as being able to “[d]iscern the military dimensions of a challenge affecting national interest,” “[a]nticipate and lead rapid adaptation and innovation,” or “execute and adapt strategy through campaigns and operations,” without a strong grasp of the fundamentals of causal analysis.⁵ As Jim Golby argues,

The U.S. military does not need or want all officers to become social scientist researchers, but applied social science can nevertheless help develop strategic thinking because it: (1) focuses on human behavior and influence; (2) develops comfort with competing theories; (3) requires creativity; and (4) uses evidence and iteration to better understand the world and adapt to change.⁶

A basic part of policymaking, strategizing, or planning at any level is attempting to anticipate the results of one’s decisions. For example, will withdrawing from the Iran nuclear deal shorten or lengthen Iran’s path to a nuclear weapon? Or will surging troops in Afghanistan lead to increased or decreased violence? Anticipating consequences is, of course, extremely difficult because national security policy is made in complicated social, strategic, military, and political environments.⁷ Any given decision pro-

duces many different results, and even in retrospect it can be difficult to establish a definitive causal link between policy and outcome. This is where social science approaches, especially those focused on causal analysis, can make significant contributions.

But while we cannot simply ignore academic social science research and methods, neither can we assume that it is immediately accessible to our students. As an extensive literature on bridging the gap between academia and the policymaking world has highlighted,⁸ the partnership between national security policymakers and social science researchers has not always been a successful one. While much of the literature focuses on what social scientists can do to make their work more accessible to practitioners, Philip Zelikow instead argues that effective policymaking relies on policymakers having adequate “software” to attack policy problems, by which he means “the way people size up problems, design actions, and implement policy.”⁹ He further argues that one reason social science research often seems irrelevant to policymakers is that “as the software of policy work has deteriorated, the people doing policy work no longer do the analysis — or articulate the questions — to seek out and use relevant knowledge, whatever its source.”¹⁰

Even in a world where academics often do not write in terms clearly accessible to a practitioner audience — and often assume a level of understanding of sophisticated methods that is a high bar even for other academics — basic literacy in the language and tools of causal analysis can still help practitioners make the most of relevant academic studies and powerful analytical tools. For the most part, national security practitioners are consumers, rather than producers, of social science research. But understanding how this research is conducted

3 “Developing Today’s Joint Officers for Tomorrow’s Ways of War: The Joint Chiefs of Staff Vision and Guidance for Professional Military Education & Talent Management,” Joint Chiefs of Staff, May 1, 2020, 2, https://www.jcs.mil/Portals/36/Documents/Doctrine/education/jcs_pme_tm_vision.pdf?ver=2020-05-15-102429-817.

4 “Developing Today’s Joint Officers for Tomorrow’s Ways of War,” 3, https://www.jcs.mil/Portals/36/Documents/Doctrine/education/jcs_pme_tm_vision.pdf?ver=2020-05-15-102429-817.

5 “Developing Today’s Joint Officers for Tomorrow’s Ways of War,” 4.

6 Jim Golby, “Want Better Strategists? Teach Social Science.” *War on the Rocks*, June 19, 2020, <https://warontherocks.com/2020/06/want-better-strategists-teach-social-science/>.

7 For more on the influences on foreign policy decision-making, see, Nikolas K. Gvosdev, Jessica D. Blankshain, and David A. Cooper, *Decision-Making in American Foreign Policy: Translating Theory Into Practice* (Cambridge, UK: Cambridge University Press, 2019).

8 For example, Alexander L. George, *Bridging the Gap: Theory and Practice in Foreign Policy* (Washington, DC: United States Institute of Peace Press, 1993); Stephen M. Walt, “The Relationship Between Theory and Policy in International Relations,” *Annual Review of Political Science* 8, no. 1 (2005): 23–48, <https://doi.org/10.1146/annurev.polisci.7.012003.104904>; Bruce W. Jentleson and Ely Ratner, “Bridging the Beltway–Ivory Tower Gap,” *International Studies Review* 13, no. 1 (March 2011): 6–11, <https://doi.org/10.1111/j.1468-2486.2010.00992.x>; James Goldgeier and Bruce Jentleson, “How to Bridge the Gap Between Policy and Scholarship,” *War on the Rocks*, June 29, 2015, <https://warontherocks.com/2015/06/how-to-bridge-the-gap-between-policy-and-scholarship/>; James Goldgeier, “A New Generation of Scholars Looks to Bridge the Gap,” *War on the Rocks*, Feb. 22, 2018, <https://warontherocks.com/2018/02/bridging-the-gap/>; Michael C. Desch, *Cult of the Irrelevant: The Waning Influence of Social Science on National Security* (Princeton, NJ: Princeton University Press, 2019).

9 Philip Zelikow, “To Regain Policy Competence: The Software of American Public Problem-Solving,” *Texas National Security Review* 2, no. 4 (August 2019): 110–27, <http://dx.doi.org/10.26153/tsw/6665>.

10 Zelikow, “To Regain Policy Competence.”

is vitally important to being an educated consumer who is able to critically evaluate and synthesize the work of others. Even if most practitioners will not find themselves in senior decision-making roles in the near future, if ever, the analytical products they produce in the form of memos and briefs do significantly shape decision-making at the highest levels. While uncertainty can never be eliminated entirely, a better understanding of the analytical tools social science can bring to bear will help policymakers and military officers successfully navigate complexity and avoid common pitfalls.

This essay attempts to make the analytical tools frequently used in social science research more “user friendly” by explaining what it means to investigate causality, discussing several primary ways analysts attempt to do so — through formal models, controlled experiments, statistical analysis, and the use of historical cases and analogies — and providing troubleshooting tips for successfully applying and evaluating these analytical methods. The aim is to provide a sort of reader's guide to sophisticated social science and policy analysis. This is not only because social science research can inform the policy process in important ways, but because critical reading is itself an important component of critical thinking. Developing and practicing this important skill will enable practitioners to make stronger contributions at all levels of policymaking.

Causality: The Holy Grail of Social Science

When civilian policymakers and military officers make decisions regarding national security policy or military strategy, they are often seeking to create (or sometimes prevent) a particular outcome. For example, a policy to increase military-to-military engagement, offer financial aid, or threaten sanctions is an effort to alter a political environment and create a political result. In such instances policymakers are making an assumption about *causality*: If I choose policy A, my preferred outcome is

more likely to occur than if I choose policy B or C.

By altering an aspect of the environment — the *cause* — one seeks to influence another aspect of the environment — the *effect*. A proposed alliance, for example, may be meant to prevent aggressive behavior on the part of potential adversaries. This was the original intent of NATO. The alliance was an effort to communicate to the Soviet Union that an attack on a single Western European member of the alliance would automatically result in a war with the other members of the alliance, including the United States. In other words, Western policymakers believed that the creation of a mutual defense treaty would cause Soviet leaders to doubt their ability to conduct a successful offensive in Western Europe, having the effect of improving deterrence vis-à-vis the Soviet Union. It is important to note that in the real world, most causes are themselves the effects of other, earlier causes. For example, one could say that the emergence of the Soviet Union as a potential hostile power caused the Western powers to form a defensive alliance.

Causality has been a key focus of the study of international relations since World War I, when many of the earliest investigations of international conflict sought to illuminate the causes of war.¹¹ After World War II, the field of security studies became more focused on “describing social behaviors as they actually occur (rather than as we might wish them to be) and explaining causal relationships among the various behaviors.”¹² As the field focused more on explaining cause and effect, it also became more “scientific” — that is, generating falsifiable hypotheses (statements that could be proven incorrect)¹³ and testing them against empirical evidence (things that happened in the real world).¹⁴ One common hypothesis you may be familiar with is the democratic peace theory, which proposes that two democratic states are less likely to go to war with one another than one democratic state and one undemocratic state.

Frequently, hypotheses will be stated (explicitly or implicitly) in terms of independent and dependent variables. The aspect of the environment the analyst

11 For a review of the early stages of causal theorizing in international relations, see, Milja Kurki, *Causation in International Relations: Reclaiming Causal Analysis* (Cambridge; New York: Cambridge University Press, 2008), chap. 1.

12 Peter D. Feaver and Erika Seeler, “Before and After Huntington: The Methodological Maturing of Civil-Military Studies,” in *American Civil-Military Relations: The Soldier and the State in a New Era*, ed. Suzanne C. Nielsen and Don M. Snider (Baltimore, MD: Johns Hopkins University Press, 2009), 72–90, 74. While acknowledging this shift, it is simultaneously important to recognize that research that simply describes the state of the world without investigating causal relationships can also make valuable contributions to knowledge.

13 A non-falsifiable hypothesis is one that cannot be disproven. A classic example has to do with the color of swans. “All swans are white” is a falsifiable hypothesis — observing a swan that is black (or any color other than white) would disprove the hypothesis. By contrast, the statement “Black swans exist” is non-falsifiable. A researcher could count white swan after white swan and still never prove definitively that black swans do not exist anywhere. To give a policymaking example, the statement, “The use of economic sanctions may allow the United States to avoid military action” is non-falsifiable. Even if one were to prove that economic sanctions have never helped to avoid military action, there is always a chance sanctions could have this effect the next time around.

14 Feaver et al., “Before and After Huntington,” 74.

believes is the cause is called the *independent variable*. It is independent because its value is assumed to be determined outside the system being studied — in other words, the analyst is not concerned with what factors affect the independent variable, but instead with the independent variable's effect on the dependent variable. In policymaking terms, the independent variable is the aspect of the environment the policymaker might seek to directly manipulate. In the democratic peace hypothesis, the independent variable is the regime type of the two states. In keeping with our example, one might note that the United States frequently promotes democratic governance around the globe.¹⁵

Changing the value of the independent variable is expected to lead to changes in the *dependent variable* — so called because its value is believed to potentially “depend” on the independent variable as well as other variables. In the democratic peace example, the dependent variable is war — have the states in question been at war with one another? Changes in the independent variable (regime type) are believed to cause changes in the dependent variable (war).¹⁶ In other words, if a totalitarian regime undergoes a revolution and becomes a democracy, other democracies should feel safer.

The difficulty, of course, comes in evaluating these causal hypotheses to determine which can be reliably used to predict real-world events, and which cannot. In a perfect world, the analyst would be able to observe two separate timelines in which the key independent variable differed. For example, to test the hypothesis that the assassination of Archduke Franz Ferdinand caused World War I, we would ideally be able to observe two parallel universes that are identical until June 28, 1914, at which point the archduke is assassinated in one universe but not the other. In technical terms, this is called evaluating the counterfactual — trying to evaluate what would have happened if one or more historical fact were changed. While counterfactuals are an extremely useful concept, it is, unfortunately, impossible for analysts to jump time streams in this manner, science fiction films notwithstanding. Instead, analysts are forced to rely on methods other than direct observation to evaluate the counterfactual, or estimate

the degree to which an action creates an effect. We discuss several of these methods for better understanding what would happen in the presence or absence of some causal factor in the next section.

Approaches to Causal Analysis

Formal Models

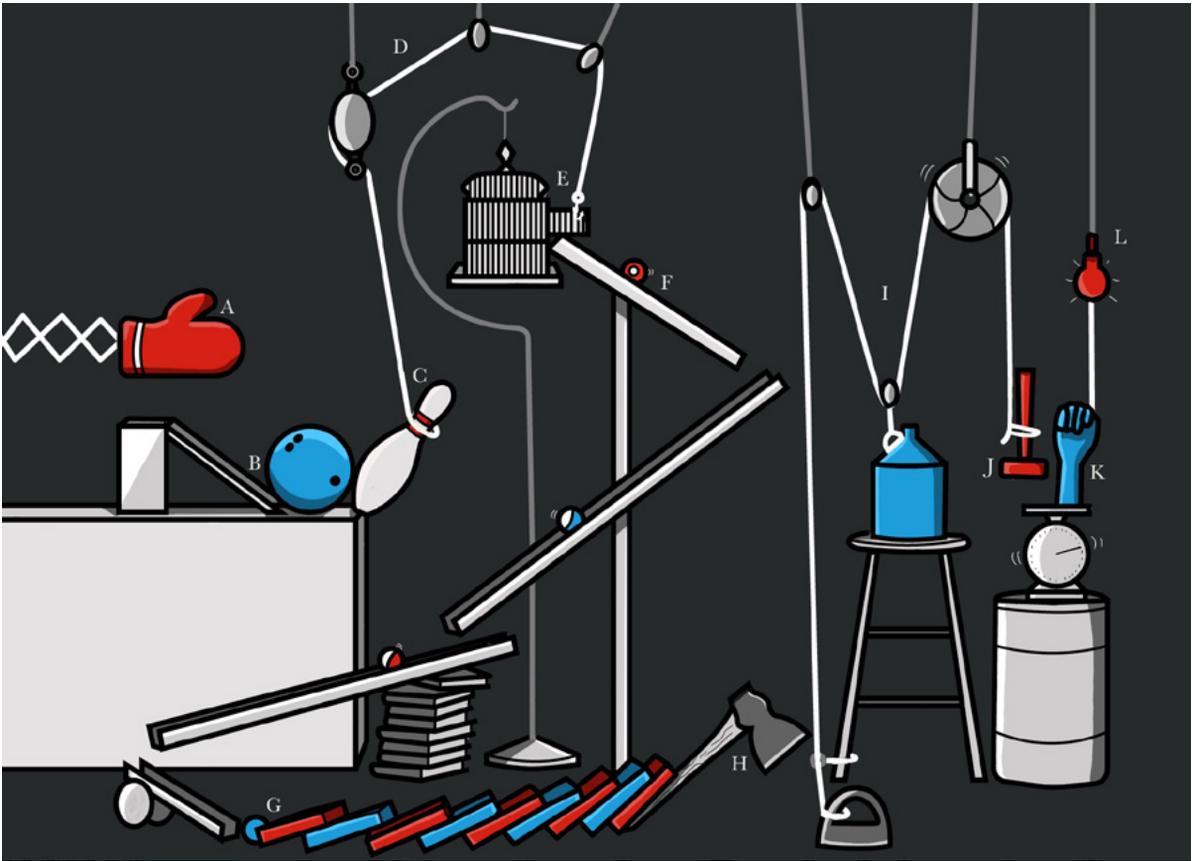
One way to better understand causal relationships is to use formal models, which are simplified representations of the world that highlight particular dynamics of importance to the analyst. The key features of a model are: 1) its explicit assumptions about the world and the actors in it, and 2) the internal logic by which it turns these assumptions into predictions, ranging from simple arithmetic to multivariable calculus. Formal models force analysts to be disciplined and transparent in their thinking about causal relationships. Models can be more or less “formal” in the extent to which they use mathematical representation and equations.

The most commonly used models in social science and policy analysis are based on rational choice theory. Rational choice theory, in its broadest form, assumes that an actor (the decision-maker), has consistent, rank-ordered preferences with regard to the different possible outcomes, and beliefs about the way their decisions or actions change the likelihood of those outcomes occurring. Rational choice theory further assumes the actor will choose the action that gives him or her the best expected outcome, or maximum expected utility.¹⁷ In such a model, an increase in the costs or decrease in the benefits expected from a particular course of action would make the decision-maker less likely to choose that option. Importantly, analysts using rational choice-based models frequently do not believe the assumptions underlying these models are realistic. States are not actually single actors who make decisions, and even people who do make decisions don't often do so by such a meticulous process. But the analyst believes the actors in their models behave “as if” these assumptions hold. In other words, the assumptions are

15 Many independent variables cannot be controlled by individual policymakers, of course. U.S. policymakers cannot wave a magic wand and turn other states into democracies. Similarly, even if leaders of authoritarian regimes determine that advances in communications technology are likely to destabilize their regime, they may not be able to stop these advances. Politicians can impact technological developments at times, but it was too late for individual political leaders to prevent communications technology from facilitating the spread of the Arab Spring after a Tunisian fruit-seller immolated himself in a public square. Other potential causal factors — such as naval deployments, diplomatic overtures, and statements of policy — may be more under the chief executive's control.

16 It is possible to discuss variation and correlation among variables without discussing causality. Some scholars argue that causation is a metaphysical notion that lies outside the realm of true knowledge. Steven Sloman, *Causal Models: How People Think About the World and Its Alternatives* (Oxford: Oxford University Press, 2005), 5–6.

17 For a more detailed discussion of rational choice theory, see, Gvosdev, Blankshain, and Cooper, *Decision-Making in American Foreign Policy*, chap. 3; and Kenneth A. Shepsle and Mark S. Bonchek, *Analyzing Politics: Rationality, Behavior, and Institutions*, 1st ed. (New York: W.W. Norton, 1997).



good enough to generate realistic predictions or outputs. Other models relax the assumptions of rational choice theory, allowing actors to have inconsistent preferences and beliefs, or to choose based on a method other than optimization.¹⁸ When these models focus on a single actor making a decision, it is generally referred to as decision theory.

Game theory refers to models that build on decision theory by adding strategic interaction between actors to the mix. In keeping with the name “game theory,” actors in these models are often called players. An individual player’s decisions are based not simply on a static view of the world, but on anticipating other players’ reactions to their own actions. Game theoretic models make assumptions about each player’s objectives or the payoffs associated with various outcomes, and about the information each player knows about the other player(s) and the structure or rules of the game. An equilibrium is a set of actions or states from which no player wants to deviate, given the other players’ anticipated responses to a deviation. The simplest game theoretic models — single-round interactions between two players — are often depicted as tables that show the players’ respective payoffs for each possible combination of actions.

The models — or games — can be “solved” by determining each player’s best response to each of the other players’ actions and whether any of these best responses form one or more equilibria. More complex games, with more than two players or multiple rounds of interaction are frequently depicted as game trees. Each node represents an actor’s decision point, with branches from that node representing possible choices. The final nodes on the tree, when there are no more decisions to be made, display each player’s payoffs from that set of possible actions. The analyst then reasons backward from these payoffs to determine likely paths the players would follow.

Perhaps the best-known example of game theory as applied to international relations is the theory of Mutually Assured Destruction (MAD). In its classic Cold War formulation, MAD involved two actors, the United States and the Soviet Union, that each ranked their own total annihilation as the worst possible outcome and believed that the other would respond to a nuclear attack with its own nuclear attack, producing total annihilation in the originator. Given these assumptions, neither country initiating an attack would produce a stable equilibrium — deviation by either party would lead to total annihila-

18 See, Gvosdev, Blankshain, and Cooper, *Decision-Making in American Foreign Policy*, chap. 4.

tion, so neither had an incentive to attack.

The United States (and NATO members) did vary at times on the role nuclear weapons would play in national and alliance defense. In 1954, Deputy Supreme Allied Commander Bernard Montgomery offered the following: “I want to make it absolutely clear that we are basing all our planning on using atomic and thermonuclear weapons in our defense. With us it is no longer ‘They may possibly be used.’ It is very definitely ‘They will be used, if we are attacked.’”¹⁹

Some American leaders, including President Dwight Eisenhower, viewed nuclear weapons to be a reliable means of national defense. But the MAD perspective on nuclear deterrence — the fear of total annihilation — arguably came to dominate Cold War thinking between the two superpowers.

In evaluating a formal model, it is important to ask whether the assumptions are reasonable for the purpose and whether the analyst seems to have faithfully followed the rules of the model’s internal logic. So, for example, in the case of MAD one could question whether either state considered some outcome to be worse than total annihilation, or doubted that the other state would actually retaliate with a devastating strike. It is important to note that such stylized models generate hypotheses — they do not test them. A model makes predictions based on its internal logic. These predictions can then be tested in the real world using some of the methods described below.

Controlled Experiments

The most straightforward and reliable way to evaluate causal claims is through controlled experiments. In an experiment, the researcher essentially applies a “treatment” (or cause) to one group of subjects but not another, and then looks for any differences in outcomes of interest between the groups. This is what medical researchers do in randomized control drug trials. The researchers randomly assign participants to the treatment and control groups, attempting to create two groups of participants who are similar in all important respects. They then give participants in the treatment group the drug being studied and participants in the control group a placebo. Any difference in outcomes between the two groups is theoretically attributable to the effect of the drug.

The double-blind randomized control trial, in

which neither researcher nor subject knows which group the subject is in, is the “gold standard” of causal analysis, but is quite difficult to administer in social science research. Social scientists are sometimes able to conduct what are called “field experiments,” taking advantage of situations where they are able to assign control and treatment groups in a natural setting, but these opportunities are rare. For example, a group of researchers used a field experiment to evaluate whether Mexico’s Seguro Popular health care program was able to successfully provide resources to low-income households. As the program was being rolled out, the researchers were able to match pairs of “health clusters,” groups of households served by a particular health care facility, and give one member of each pair a treatment that involved encouraging individuals to enroll in the health program and providing resources for improved health facilities.²⁰

One area of social science research where experiments are more common is public opinion research. Analysts use survey experiments to observe how respondents react to various treatments incorporated in the survey design. One set of respondents gets a particular version of the survey, while others get a different version or versions that vary the questions or framing. In one study, for example, researchers asked American survey respondents whether they would support the United States engaging in a hypothetical military action. Some respondents (the control group) were given no additional information about the U.S. armed forces that would engage in this action. Respondents randomly selected into the treatment group, however, were told that policymakers were considering reintroducing a draft and that this shift would precede the hypothetical military action. The researchers were thus able to estimate whether the reintroduction of a draft would affect support for military action.²¹

Frequently, however, the realities of human behavior, not to mention ethical concerns, prevent social science analysts from running experiments to create ideal circumstances for causal analysis. With respect to government policymaking, in particular, policymakers frequently only get one shot — they cannot simply adjust their treatment and try again, as medical researchers do. Instead, policy analysts frequently try to find quasi-experiments, where existing features of society or the randomness of nature have essentially randomly separated peo-

19 Lawrence Freedman, *The Evolution of Nuclear Strategy* (New York: St. Martin’s Press, 1981), 84.

20 Gary King, et al., “Public Policy for the Poor? A Randomised Assessment of the Mexican Universal Health Insurance Programme,” *The Lancet* 373, no. 9673 (April 2009): 1447–54, [https://doi.org/10.1016/S0140-6736\(09\)60239-7](https://doi.org/10.1016/S0140-6736(09)60239-7).

21 Michael C. Horowitz and Matthew S. Levendusky, “Drafting Support for War: Conscripted and Mass Support for Warfare,” *Journal of Politics* 73, no. 2 (April 2011): 524–34, <https://doi.org/10.1017/S0022381611000119>.

ple into groups for which the independent variable varies, but other important characteristics do not. Take, for example, a quasi-experiment that sought to determine the effect of independent media on the 1999 Russian parliamentary elections. The one national television channel that was independent from the Russian government was accessible to only part of the population. The researchers argue that access to it was effectively random once they controlled for observable factors such as population size and urban status (see the next section on statistical analysis). They found that access to the independent channel decreased the government's vote share and increased the vote share for opposition parties.²² While this was not a true controlled experiment, it is about as close as one can come in the real world of policy analytics.

Military operations offer more opportunities for actual real-world hypothesis testing, in some respects. The physical movement and interaction of forces can reveal facts that are observable on the ground (or in the air or sea or cyberspace). In training and in wargaming, actions can be repeated under different circumstances, or different actions can be tried in the same circumstances, to attempt to discern a relationship between cause and effect. Other critical causal relationships — such as deterrence — are more difficult to evaluate in this fashion since the results are not purely physical and thus can be more difficult to observe. Deterrence played a major role in the conduct (if one could call it that) of the Cold War, but it is almost impossible to identify which actual military steps helped generate a deterrent effect.

So how is a practitioner to evaluate experimental or quasi-experimental evidence for causal relationships? As discussed above, key to the effectiveness of an experiment is the random assignment of “participants” to control and treatment groups. It is important to consider whether these assignments were effectively random. Could the control and treatment groups vary in a systematic way that biases the results? Also important is the “operationalization” of the treatments and the outcome: Do the treatments actually create the conditions they were supposed to create? Is the dependent variable truly measuring the outcome of interest? Another primary concern with social science experiments is what academics call “generalizability” or “external validity.” Would the results likely hold outside of the precise experimental conditions studied? If the subjects were aware that they were participating in an experiment

(true experiments frequently require voluntary consent), might they have behaved differently than they would have in a “normal” setting?

Statistical Analysis

As another option for causal analysis, social scientists sometimes turn to what they call “large-N data.” Rather than investigating the details of a specific historical case, or cases, the analyst collects more limited information on a large number of cases, which are often called “observations.” The goal is to use the large amount of data to control for variation among the observations. For example, to test the democratic peace hypothesis, referenced above, one might look at a large dataset of countries over time to determine whether pairs of democratic countries are, in fact, less likely to go to war with one another than are pairs of countries that include a non-democracy. The large dataset would allow the analyst to control for so-called “confounding variables,” such as whether each pair of countries shares a land border. There are enough pairs of countries that the analyst can use statistical techniques to determine whether the democratic peace relationship holds for both countries that share a border and those that do not.

If you've ever taken a statistics course, you may have heard the statement “correlation does not mean causation,” but what does that actually mean? What, other than a cause and effect relationship, might explain a correlation (relationship between two variables) in large-N data? There are three possibilities:

Selection Effects: One possibility is that an unintended force is “choosing” observations for inclusion in the dataset. During World War II, for example, the civilian Statistical Research Group was tasked with improving the design of fighter aircraft. Initially, the military researchers conducted a survey of combat damage on returning aircraft, believing that the parts of the aircraft that were disproportionately struck by enemy fire should receive better armor. A statistically trained member of the staff realized this was a mistake. The survey was of a non-random sample of fighters — they were only looking at the aircraft that had returned from combat. A better starting point would be to assume that the areas of the returning combat aircraft that were *not* damaged are exactly the parts of the plane that should receive additional protection. Combat damage to those areas was more critical, since planes struck in those areas were likely fatally damaged and did not return from the mis-

22 Ruben Enikolopov, Maria Petrova, and Ekaterina Zhuravskaya. “Media and Political Persuasion: Evidence from Russia,” *American Economic Review* 101, no. 7 (2011): 3253–85, <https://www.jstor.org/stable/41408737>.

sion. This latter approach was eventually adopted by both the Navy and the Air Force.²³ In this case, the research design was out of the analysts' hands — the only planes available to examine were those that survived. But researchers sometimes inadvertently introduce selection effects in their results by choosing to include only “successful” cases, or cases where the expected outcome occurred. This is sometimes called “selecting on the dependent variable.” Consider the following statement: “Study finds that 80 percent of successful senate campaigns raised over \$X million.” It is tempting to infer that high spending leads to campaign success, but without knowing how much losing campaigns spent, we cannot make this inference.

Omitted Variable Bias: Another possibility is that some other, unidentified variable is causing variation in both the independent and dependent variables. Consider, for example, the statistic that American military officers disproportionately identify with the Republican Party when compared to the overall U.S. population.²⁴ One might note, however, that American military officers are also disproportionately white males, compared to the overall population. And white males are more likely to be Republicans than are individuals in many other demographic groups. So it could be that omitted variables — sex and race — are driving the relationship between officer status and partisan affiliation, rather than any causal relationship between military service and party identification.

Reverse Causality: It is also possible that the analyst has misidentified the causal relationship between the independent and dependent variables — that is, changes in the dependent variable cause changes in the independent variable, rather than vice versa. Consider exit polling at elections. Say a particular poll found a high correlation between voters who say that immigration is the most important issue to them, and voters who voted for Donald Trump. One might infer that those voters chose Trump *because* they care about immigration and Trump highlights this issue in his stump speeches. It is also plausible, however, that these individuals say they care about immigration *because* they support Trump and are therefore frequently exposed to his positions on immigration.

There are, of course, many other important considerations in evaluating a statistical model. If the statistical model chosen is not a good fit for the

data, or there is significant error in measuring variables, the model's results may be biased. Data available may also not be an exact match for the concept the analyst is trying to represent in the model, or the analyst may need to attempt to quantify inherently qualitative variables (for example, the degree to which a state is a democracy). These types of issues can be more difficult for a non-expert to detect. A more casual reader frequently has to put some amount of faith in the peer-review process by which many academic articles are published and trust that reviewers and editors with appropriate expertise will have identified such issues.

In many cases, a practitioner (or even an academic) reading a statistical analysis won't necessarily be familiar with the details of the particular regression model or statistical technique being employed. But that does not mean the reader simply needs to take the author's word that the results are convincing. An educated consumer of these types of analyses can still consider whether the variables, as defined by the author, measure what they are intended to measure; whether any causal claims might be confounded by selection effects, omitted variables, or reverse causality; and whether the author has successfully corrected or accounted for such possible confounders. And again, readers should consider external validity: Is there reason to believe the relationship captured might be specific to the study's sample or to particular contexts? As a reader, it is also important to consider the source — are you reading a peer-reviewed scholarly article? A report from a well-known think tank? A journalist's summary of an unpublished study? While there are certainly bad studies published in academic journals, and good work done by data journalists, the answer should influence your level of trust or skepticism toward the findings.

Counterfactuals: Historical Cases and Analogical Reasoning

Yet another method social scientists and policy analysts use to evaluate causal relationships is investigating historical cases and analogies. Analogies, or comparisons between a current situation and examples of similar foreign policy problems in the past, are commonplace in national security debates and policymaking. Analogies are probably unavoidable in national security decision-making, since one

23 W. Allen Wallis. "The Statistical Research Group, 1942-1945: Rejoinder," *Journal of American Statistical Association* 75, no. 370 (June 1980): 334-35, <https://www.jstor.org/stable/2287454>.

24 See, for example, Hugh Liebert and James Golby, "Midlife Crisis? The All-Volunteer Force at 40," *Armed Forces & Society* 43, no. 1 (January 2017): 115-38, <https://doi.org/10.1177/0095327X16641430>.

cannot make a policy decision without making some reference to history.²⁵ However, analogies can harm the interests of the policymaker as well, if they promote a misunderstanding of the policy options.

When social scientists use historical case studies — sometimes called “small-N analysis” — they choose cases carefully depending on the purpose of the analysis.²⁶ The cases can be used to develop hypotheses, to illustrate the plausibility of causal mechanisms, or to test hypotheses. For example, comparing two cases that are as similar as possible in all but one respect can help to test hypotheses about the consequences of that difference. In one study, Deborah Avant compares American and British systems of military oversight — and particularly their respective experiences in Vietnam and the Boer Wars and the Malayan Emergency — to argue that divisions in the civilian government make the military less responsive to civilian attempts to encourage military innovation.²⁷ Alternately, instead of comparing cases, one particular case can be examined in detail using a method called process-tracing, which helps to evaluate the plausibility of particular causal mechanisms by looking for them in action. Importantly, the same case(s) cannot be used for both theory-development and theory-testing, because this can bias the analysis and reduce generalizability.

In policymaking, analysts often invoke particular historical analogies to predict what is likely to happen in a developing situation. This can be dangerous, as Yuen Foong Khong highlights in his analysis of the internal deliberations of the administration of Lyndon Johnson in 1966. Khong describes a May 6 memo in which Walt Rostow, then national security adviser (but also an academic economist), offered a comparison between the vulnerability of Nazi Germany's supply chain of petroleum, oil, and lubricants (POL) during World War II, and the POL vulnerability of North Vietnam.²⁸

Rostow pointed out that the Allies' assessment of Germany's logistics during World War II had overestimated Germany's ability to reallocate POL supplies from civilian uses to military uses. As a consequence, as soon as systematic attacks on Germany's oil supplies began, there was an immediate impact on Germany's military capability — in fact, the bombing had

a far greater impact than had been anticipated.

In 1966, American assessments of the vulnerabilities of North Vietnam's oil logistics were similarly pessimistic that a sustained air campaign would have a meaningful impact on Hanoi's ability to supply guerrillas in the South. Rostow used his intimate familiarity with the example of Nazi Germany to argue that American assessments of North Vietnam were similarly off the mark:

We used then [in 1944] with respect to the Germans exactly the same analytical methods we are now applying to North Vietnam ... Assuming that they [the Germans] would and could cushion front line military requirements, we told our seniors that attacks on oil would be considerably cushioned and delayed in their impact on the military situation in the field.

We were wrong. From the moment that serious and systematic oil attacks started, [the Germans'] front line fighter strength and tank mobility were affected. ...

*With the understanding that simple analogies are dangerous, I nevertheless feel it is quite possible the military effects of a systematic and sustained bombing of POL in North Vietnam may be more prompt and direct than conventional intelligence would suggest.*²⁹

As events in 1966 proved, Rostow was wrong. Johnson went ahead with the bombing that Rostow recommended, but North Vietnam's military support for the insurgency was not dramatically impacted by attacks on oil production and transport.³⁰

Khong notes that Rostow explicitly acknowledges the dangers associated with analogical reasoning and yet still falls into the trap.³¹ This is an important caveat, one that can be applied to the full spectrum of policy analysis: Recognition of the possibility of certain types of errors in reasoning is no guarantee whatsoever that one can avoid committing those same errors. In fact, there is no guarantee that deeper understanding of the potential pitfalls of human reasoning, in it-

25 For much more on the use of history in policymaking, see, Richard E. Neustadt and Ernest R. May, *Thinking in Time: The Uses of History for Decision-Makers* (New York: Free Press, 1988).

26 Alexander L. George and Andrew Bennett, *Case Studies and Theory Development in the Social Sciences* (Cambridge, MA: MIT Press, 2005).

27 Deborah D. Avant, *Political Institutions and Military Change: Lessons from Peripheral Wars* (Ithaca, NY: Cornell University Press, 1994).

28 This example is discussed in Yuen Foong Khong, *Analogies at War: Korea, Munich, Dien Bien Phu, and the Vietnam Decisions of 1965* (Princeton, NJ: Princeton University Press, 1992), 209–11.

29 Walt Rostow, "Memo to the President," May 6, 1966, excerpted from Khong, *Analogies at War*, 209–10, emphasis Khong's.

30 Khong, *Analogies at War*, 210.

31 Khong, *Analogies at War*, 211.

self, leads to better decision-making. There is also a fine line between being aware of the perils of analytical reasoning and becoming frozen with self-doubt and over-critical reflection. Be aware of the human capacity for error, but do not let that awareness lead to cognitive immobilization.

This may be unsatisfying advice, but there are rarely by-the-numbers solutions in the arena of reasoning and analysis. There is always opportunity to choose theories and evidence that support one's prior conclusions. Rostow, for example, always adhered to his mistaken assessment of the overall situation in Vietnam, claiming until his death that America's involvement in Vietnam was a success because it allowed Southeast Asia time to develop economically, thereby preventing the fall of the regional dominoes that the Americans feared in the late 1950s and early 1960s.³²

Historical analogies may help the analyst to "reason out" a counterfactual. Arguably, the best counterfactuals (in terms of how reliably they can be applied to current decisions and unfolding situations) are those that 1) contain only small changes from actual historical events, and so remain as close to history as possible; 2) involve plausible or imaginable deviations from the historical record (such as, "What if Al Gore had won the 2000 presidential election?"); and 3) are open to the possibility that historical currents could push the counterfactual environment back toward the original, actual outcome.

Khong has developed an analogical explanation framework, which "suggests that analogies are cognitive devices that 'help' policymakers perform six diagnostic tasks central to political decision-making"³³:

1. *Define the Situation.* By offering a comparison from the past, one can help cast a different light on the policy environment that is currently under consideration. Is a policy choice similar to Neville Chamberlain's decision to be conciliatory to Adolf Hitler at Munich in 1938, in that it would be a mistake to miss an opportunity to nip a growing threat in the bud? Or is it similar to the American decision to escalate in Vietnam, in that it would be a mistake to become stuck in a disastrous quagmire?
2. *Assess the Stakes.* How serious are the potential outcomes of the current situation? What is the range of risks to be considered? Is the situation more like the Cuban Missile Crisis of 1962 or

the Crimean Crisis of 2014? By making historical comparisons, one can potentially assess the possible repercussions of a policy choice.

3. *Provide Prescriptions.* If historical comparisons are appropriate, they may suggest a range of prescriptions, or policy options. Options considered in past crises may be applicable to new situations.
4. *Predict Likelihood of Success.* Historical comparisons can aid in determining the likelihood of success of particular policy options, or the likelihood of various outcomes. This is perhaps the most perilous use of analogies, since it applies directly to the core question that faces that policymaker: What is the probability a policy will achieve the desired result?
5. *Evaluate Moral Rightness.* Historical comparisons can help construct a moral framework around a policy choice. During the Cuban Missile Crisis, some of the advisers to President John F. Kennedy recommended a surprise military strike on Cuba to eliminate the missiles. As this option was discussed (and ultimately rejected), Robert Kennedy supposedly slipped a note to his brother that read, "Now I know how Tojo felt when he was planning Pearl Harbor."³⁴ For men who had lived during World War II, and would have clearly recalled the nation's rage and moral indignation following the attack on Pearl Harbor, this would have been a telling comparison.
6. *Warn About Dangers.* Analogies can potentially warn of the dangers that may be unwittingly embraced by pursuing particular policy options. For example, Gen. Douglas MacArthur's actions as he pursued retreating North Korean forces during the earlier stages of the Korean War increased the likelihood of Chinese involvement in the conflict. Do current policy choices risk an inadvertent escalation with a third party, or other dangers? By exploring historical examples of how policies went wrong, one can better avoid similar pitfalls in the future.

In addition to this framework laid out by Khong, national security practitioners should keep in mind that analogies are often used as dramatic comparisons to strengthen an argument, rather than as coldly logical comparisons meant to illuminate a line of reasoning. When

32 "Walt Rostow," *The Economist*, Feb. 20, 2003, <https://www.economist.com/obituary/2003/02/20/walt-rostow>.

33 Khong, *Analogies at War*, 10. The framework is introduced in chapter 1 and explained in detail in chapter 2. Explanations and illustrative examples have been added to Khong's list of central diagnostic tasks.

34 Robert F. Kennedy, *Thirteen Days: A Memoir of the Cuban Missile Crisis* (New York: W.W. Norton, 1999), 31; see also, Ernest R. May and Philip D. Zelikow, eds., *The Kennedy Tapes: Inside the White House During the Cuban Missile Crisis*, Concise ed. (New York: Norton, 2002).

historical analogies are used in policy debates, it is usually to suggest that if a policymaker doesn't do X, it will lead to another Munich/Pearl Harbor/September 11/Vietnam. The dramatic images and emotions that are summoned by making reference to iconic past events are used to sway the listener — and, at times, to reassure the person making that reference that their proposed policy decision is the right one.

It is certainly impossible to resolve all dimensions of causal complexity prior to deploying a policy, but being alert to the complex nature of causality in national security policy is essential.

Lastly, practitioners should be alert to the silent analogies, the events that did not occur. The examples in history of the wars and policy disasters that were avoided — due to careful reasoning, risk aversion, or a rigorously observed policy process — are little reflected on in history, since they were non-events. Historians expend far more effort answering the question, “Why did history unfold the way it did?” than investigating “What is the version of history that didn't take place?” Non-events are notoriously difficult to investigate and explain, since they often leave a far less precise and less traceable historical record. Non-events also receive far less attention from the electorate. As former Rep. Barney Frank, who was chairman of the House Financial Services Committee and one of the architects of the plan that averted a severe economic downturn following the financial crisis of 2008, noted, “You get no credit for disaster averted or damage minimized.”³⁵

But these silent analogies must not be overlooked. Many are policy triumphs in their own right — occasions when diplomatic or military measures prevented a setback or a disaster. Or they may be examples of how individuals recognized the potential for a policy pitfall and avoided it. If President Barack Obama had launched

a strike on Syria in September of 2013, would the United States have been set on course for a deeper and doomed engagement in the Middle East? While his decision to approach Congress for an authorization of force was unexpected and was seen by many as an example of weakness,³⁶ it may have been part of a causal chain of events that avoided deeper engagement in a civil war that arguably could not be resolved by American involvement.

In short, when considering an analyst's use of historical cases or analogies to investigate a causal relationship, it is important to consider whether they were chosen in an appropriate way. It is also important to consider whether the case or analogy seems to be rendered accurately — i.e., are key features of the case being twisted or glossed over to make it a better fit for the causal argument? Finally, external validity is a key consideration. If you are convinced a causal relationship held in one or more cases, is it likely to hold in other cases? If so, what conditions might those cases need to fit?

Troubleshooting Causal Analysis in National Security

To conclude, below are some final considerations for reading, evaluating, and applying sophisticated analytical work in a national security policymaking context³⁷:

Know the Goal, Which May Not Be “Success,” Narrowly Defined

While causal reasoning is important to understanding the likely implications of a policy decision, it is not the only thing that matters. Policy choices are often dependent on factors other than likelihood of success (such as domestic politics). One may have reason to believe that a policy will “work,” but choose not to pursue it for moral, fiscal, or other reasons. In other words, even when you are relatively confident in a causal relationship between policy X and outcome Y, policy X might also affect outcomes A, B, and C, which are just as important to senior policymakers.

35 Andrew Ross Sorkin, “President Obama Weighs His Economic Legacy,” *New York Times Magazine*, April 28, 2016, <https://www.nytimes.com/2016/05/01/magazine/president-obama-weighs-his-economic-legacy.html>.

36 Peter Baker and Jonathan Weisman, “Obama Seeks Approval by Congress for Strike in Syria,” *New York Times*, Aug. 31, 2013, <https://www.nytimes.com/2013/09/01/world/middleeast/syria.html>.

37 Portions of this discussion draw on Andrew L. Stigler, “Assessing Causality in a Complex Security Environment,” *Joint Forces Quarterly* 76, no. 1 (January 2015), <https://ndupress.ndu.edu/JFQ/Joint-Force-Quarterly-76/Article/577586/assessing-causality-in-a-complex-security-environment/>.



Make Causal Reasoning as Explicit as Possible

There are countless examples of major policy decisions that relied on casual and dubious causal reasoning. For example, consider the expectation that a successful imposition of a democratic government in Iraq following the 2003 U.S. invasion would create a “beacon of democracy” in the region that would then lead to other democratic transformations.³⁸ There is a spectacular array of causes and effects at work in this proposal to invade a country, transform its government, and thereby alter the political trajectory of an entire region. What was the likelihood an invasion would lead to democratic transition taking place in Iraq? How did this likelihood depend on the specifics of the invasion? If such a transition did take place, would the emergence of a stable democracy in Iraq be viewed by other states in the region as an appealing alternative form of government, or as a case of an outside power imposing a system by force? What possible setbacks or other events could reduce or eliminate the “beacon effect” of an Iraqi democracy?

At the same time, there has never been a major policy decision that resolved every causal conundrum before the policy was put into action. Even though years of thought and preparation had preceded the invasion of Europe on June 6, 1944, Gen. Dwight Eisenhower was in no way certain of the outcome. In fact, he famously kept two speeches in his jacket pockets that day, one written in anticipation of success and one in case of failure. It is certainly impossible to resolve all dimensions of causal complexity prior to deploying a policy, but being alert to the complex nature of causality in national security policy is essential.

Assess All the Available Evidence and All the Available Arguments

One should do as thorough a job as possible of assessing the available evidence about whether a causal relationship will or will not hold. During the years preceding the Vietnam War, American policymakers feared that a communist takeover of South Vietnam and unification of the country would lead to a series of Vietnam’s neighbors succumbing to communist influence as well — a domino effect.

But what evidence was there that national borders would prove so porous to communist insurgencies, and that Vietnamese nationalists would

seek to spread communism’s influence in the region instead of consolidating their control of the country? Wouldn’t nationalist movements in Cambodia and Laos work against such a contagion effect? Aren’t there few historical examples of such a social contagion? Fuller consideration of these countervailing arguments could have undercut fears of such a political domino effect taking place after the unification of Vietnam.

Of course, it is never possible to fully assess all the available information relevant to any given situation. As former Secretary of State and former Chairman of the Joint Chiefs of Staff Colin Powell once suggested in regards to how much data you need at your fingertips to make decisions, an “80% solution” is often good enough.³⁹ What Powell had in mind was that a leader who waits until all relevant information has been collected and assessed has almost always missed the opportunity to deploy a policy at a time of maximum potential effectiveness.

Similarly, evidence that is consistent with one hypothesized causal relationship may also be consistent with other possible causal relationships. To assess whether a particular piece of evidence actually supports one causal relationship over another, it helps to make all the alternatives explicit and evaluate whether each piece of evidence supports all, some, or none of the possible arguments.

The Practical Importance of Causal Analysis

We conclude this primer with a vignette to illustrate the dangers of failing to critically evaluate evidence in applying causal reasoning. In early 1999, as Yugoslav Serbs escalated their attacks on Albanian Muslims in Kosovo, the United States faced a quandary. Throughout the 1990s, America had been reluctant to commit large-scale ground forces to attempt to counteract the humanitarian disasters that unfolded during the dissolution of Yugoslavia. At the same time, the long march of atrocities in the Balkans was of deep concern to the United States and NATO.

As Kosovo became the new flashpoint in the region, policymakers again considered their options. At this time, Ambassador Richard Holbrooke’s account of

38 This argument was one of several that were offered in support of the 2003 invasion, with others including the now-infamous argument that Iraq had weapons of mass destruction. There is room to debate which arguments were predominately part of the political marketing campaign of the policy decision and which truly drove the George W. Bush administration’s decision-making.

39 Clark D. Stuart II, *Battlefield to Boardroom: Lessons Learned from U.S. Navy SEALs* (Bloomington, IN: Trafford, 2006), <https://www.overdrive.com/search?q=5C6FED79-3118-4ADB-8014-20CBBEE56F4F>, 116.

the diplomatic efforts that led to the successful Dayton accords in 1995 had just been published.⁴⁰ In it, Holbrooke offered what many accepted as conventional wisdom: The use of NATO airpower against the Yugoslav Serbs played a key role in coaxing Slobodan Milosevic to change his mind about seeking a political resolution at Dayton. Many of those who had worked with Holbrooke and sympathized with his perspective — including then-Secretary of State Madeleine Albright — were still in government. Key policymakers reasoned that if Milosevic reversed course in 1995 after NATO airpower was unleashed, he would probably do so again in Kosovo.

In March of 1999, NATO launched Operation Allied Force, a limited air campaign aimed at coercing Milosevic to end Serbian depredations in Kosovo. But Milosevic's response was not what the Western allies anticipated. Almost as soon as bombing commenced, the Serbs in Kosovo escalated their attacks, displacing hundreds of thousands of Kosovar Albanians. Now NATO found itself in an undesired position: The alliance had inadvertently committed itself to reversing a worsening situation, but with no strategy for guaranteeing a positive outcome.⁴¹ Instead of a short bombing campaign, the aerial assault continued for 78 days, and the political outcome was almost always in doubt throughout the conflict.

Ultimately, the NATO campaign succeeded in removing Yugoslav forces from Kosovo, but the sequence of events had not followed the allies' expectations. Milosevic had certainly not repeated the same pattern he had followed in 1995. To the chagrin of President Bill Clinton and others, the analogy of 1995 that had played a role in convincing cautious leaders to initially approve the attack proved to be a flawed comparison. This example highlights the difficulty of operating in a complex policy environment where the link between action and outcome is not always clear. It appears policymakers failed to correctly anticipate the effects of their chosen policy — in this case by misapplying analogical reasoning.

While analytical tools are not guarantors of policy success, this article has offered a range of analytical concepts that can be of use to national security practitioners, military and civilian alike. While the vast majority of practitioners will not be designing major U.S. campaigns or strategies on their own, their analytical products shape and drive the policymaking process in important ways. An aware-

ness of causal processes, complexity, analogical reasoning, and elementary statistical pitfalls can considerably improve the underlying analysis that drives national security decision-making. As Golby notes, "the broader interactive and adaptive approach that social scientists use relies on the same fundamental methods and concepts that strategic leaders must replicate, usually more quickly, in practice."⁴² There is almost never any danger in rendering analytical processes as explicitly as possible, since explicitness can only reveal biases. To paraphrase Robert Fogel, everyone has analytical biases and the only alternative to open concession of these biases is to conceal them.⁴³

The methodological techniques described above can prove useful in unexpected places. Alertness to analytical and methodological issues is a critical facet of policy development in national security. While most national security practitioners will not be conducting original research, understanding analytical concepts is crucial to being able to evaluate others' research and one's own analysis of primary and secondary sources. The fact that many of these potential pitfalls and opportunities go unmentioned does not mean that they are not critically important. A more solid grounding in these aspects of policy analysis can greatly improve one's contribution to any national security decision-making process. 📌

Jessica D. Blankshain is an assistant professor of national security affairs at the U.S. Naval War College in Newport, RI. She writes on U.S. civil-military relations and foreign policy decision-making, and is the co-author of *Decision-Making in American Foreign Policy: Translating Theory into Practice* (Cambridge University Press, 2019).

Andrew L. Stigler is an associate professor of national security affairs at the U.S. Naval War College in Newport, RI. He is the author of *Governing the Military* (New York: Routledge, 2019).

This article reflects the personal views of the authors. It does not represent the views of the U.S. government, Department of the Navy, or U.S. Naval War College.

Photo: Kurt. S.

40 Richard Holbrooke, *To End a War* (New York: Modern Library, 1999).

41 See, for example, Jane Perlez, "Crisis in the Balkans: News Analysis; 3 Options for Washington, All with Major Risks," *New York Times*, May 21, 1999, <https://www.nytimes.com/1999/05/21/world/crisis-balkans-analysis-3-options-for-washington-all-with-major-risks.html>.

42 Golby, "Want Better Strategists?"

43 Philip E. Tetlock and Aaron Belkin, eds., *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives* (Princeton, NJ: Princeton University Press, 1996), 4.