

The Influence of Psychological Factors in the Search for Strategic Stability

Rose McDermott



Many of the leaders of new nuclear powers sit atop regimes that are personalistic in nature, and thus less constrained by institutions or public opinion than more democratic regimes. Such authoritarian leaders have more freedom to allow their psychological proclivities to influence their decisions and behavior. These tendencies include four psychological biases that influence decision-making: overconfidence, the planning fallacy, the illusion of validity, and the prominence effect. The ubiquity of these biases challenges our existing assumptions about rationality and the nature of strategic stability. The introduction of new and potentially more dangerous weapons—including the use of artificial intelligence or hypersonic missiles—heightens the risk of misperception, miscalculation, increased time pressure, and other factors exacerbating the threats confronting the world’s nuclear powers. Lessons from both cognitive and evolutionary psychology are discussed, as well as some policy recommendations to help improve decision-making under conditions of dominant leadership styles.

Many of the leaders of new nuclear powers, or states seeking such capacity, sit atop regimes that are more personalistic in nature—and thus less constrained by institutional and state structures or public opinion—than more democratic regimes. As a result, such authoritarian leaders have more freedom to allow their psychological proclivities to influence their decisions and behavior.¹ These factors challenge our existing assumptions about the nature of strategic stability and nuclear deterrence; personalistic leaders fall prey to the kinds of psychological biases that have been demonstrated to lead to suboptimal decisions. The introduction of new and potentially more dangerous weapons—including the use of artificial intelligence (AI) or hypersonic missiles—heightens the risk of misperception, miscalculation, increased time pressure, and other factors exacerbating the threats confronting the world’s nuclear powers, of which personalist and autocratic leaders already pose a heightened risk.

Nuclear deterrence and strategic stability have always rested on fundamental assumptions about

the nature of human psychology that are dubious at best and demonstrably empirically false at worst. Rationality, for example, represents one of the beliefs that have little, or at least restricted, basis in reality. And yet the entire fate of the planet depends, in essence, on our collective delusion that pivotal actors, no matter how selfish or narcissistic, remain fundamentally rational and intent on preserving their states and, even more importantly for personalist leaders, their own lives. Like most judgmental biases, this perception may be mostly accurate most of the time. Some of the time, however, as in the case of a suicide bomber, it may be wrong to assume people prize their lives over their ideals. In other words, the entire notion of rationality encompasses systematic empirical flaws that produce predictable biases in decision-making that can undermine the very foundations of stability and deterrence. Many of these biases are only enhanced with the use of emerging technologies such as AI because such models, built on the basis of human input, often serve to magnify human biases, as has been the case with sexist and misogynistic content.²

1 Jessica Weeks, "Strongmen and Straw Men: Authoritarian Regimes and the Initiation of International Conflict," *American Political Science Review* 106, no. 2 (2012): 326–47; Rose McDermott, "Psychology, Leaders, and New Deterrence Dilemmas," in *The Fragile Balance of Terror: Deterrence in the New Nuclear Age*, ed. Vipin Narang and Scott Sagan (Cornell University Press, 2022).

2 Andrew Hundt, William Agnew, Vicky Zeng, Severin Kacianka, and Matthew Gombolay, "Robots Enact Malignant Stereotypes," *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (June 2022): 743–56.

As increasing, if not novel, threats enter the domain of nuclear politics, concerns about the reliability and dependability of the American deterrence posture heighten. In particular, the Trump administration's terminating the Iran nuclear deal during his first term and withdrawing support for Ukraine during the current term, among many other instances, increase the prospects for enhanced proliferation. Moreover, Vladimir Putin's nuclear saber-rattling increases the risk of nuclear weapons actually being used in conflict.

In sum, conventional wisdom around the stability of nuclear weapons rests on the notion that nations, and their leaders, are rational actors who will be deterred from the use of nuclear weapons by the assurance that such action could result in retaliation that would destroy their own interests and their very lives. But this theory rests on an enormous number of tenuous assumptions. For example, what if a narcissistic leader were facing a painful terminal illness? Would the prospect of taking everyone else down with them prove a deterrent to action? Unlikely. What if the leader, out of arrogance or narcissism, does not believe the other side would be able to retaliate? Are personalist leaders overly prone to "techno-optimism," such that they will misjudge the risks highlighted in other articles in this Roundtable? Such factors may have seemed less likely in the past but appear increasingly probable as the number of personalistic regimes possessing or seeking weapons of mass destruction proliferates. Some people may assume that because deterrence has held for seventy-plus years, it will remain stable indefinitely. But such beliefs do not consider the increase in simple mistakes or intentional action that occurs as more nuclear states enter the system, or more personalist leaders gain access to nuclear weapons. Just because deterrence held in the past does not guarantee its maintenance in the future. Indeed, as first-person memory of the horrors of Hiroshima and Nagasaki fades, the salient visual images of destruction become less available as an emotional safeguard against such action.

Moreover, these factors do not operate in isolation. The pace of innovation in emerging technologies, ingrained human psychological biases involving perceptions of threat and desires for vengeance, and an increase in personalist regime types interact to produce new challenges. The collective impact of such factors on outcomes of interest, including strategic stability, constitute more of a risk than any of these forces in isolation. Their interaction alone raises the possibility of deterrence failure in a way that any one of them in isolation may not.

This article proceeds as follows. The first section discusses four judgmental biases that can affect decision-making around nuclear deterrence: overconfidence, the planning fallacy, the illusion of validity, and the prominence effect. For each, I review findings from the study of psychology, connect them to the theory and strategy of nuclear deterrence, and point out how emerging technologies may exacerbate these risks. The second section of the article integrates two psychological traditions—Kahneman and Tversky's cognitive revolution and evolutionary psychology, respectively—that are often, albeit unnecessarily, viewed as opposing in the study of conflict. The third section offers policy implications for how to treat decision-making as a skill that can be both learned and improved.

Important and Common Psychological Biases in Decision-Making

Many judgmental biases could potentially affect decision-making in general, with profound implications for the stability of nuclear deterrence, and emerging technologies (such as AI) could heighten the impact of a number of these. Following a brief discussion of the study of decision-making under conditions of risk, this section explores the four judgmental biases noted above, whose recognition can help improve our understanding of the way that human psychological architecture may not be well suited for understanding the enormous destructive potential of weapons in the nuclear age, or the impact of emerging technologies on the prospects for human survival. The rapid pace of such technological change occurring much faster than any true reckoning with the ethical implications or the potential negative fallout of such innovations, as well as the inability or unwillingness to regulate the industry or its growth, only serve to exacerbate the risks of catastrophic violence. Alone, and in concert, each of the psychological biases discussed here can influence how leaders, especially personalistic ones, can render incorrect judgments and make destructive decisions without realizing the bias inherent in their choices.

Perhaps the most well-known work examining the effect of judgment under uncertainty and decision-making under conditions of risk is Nobel Prize-winner Daniel Kahneman's *Thinking, Fast and Slow*, which provides a broad overview of this entire field of inquiry.³ In this work, also discussed in Herbert Lin's contribution to this Roundtable, Kahneman

3 Daniel Kahneman, *Thinking, Fast and Slow* (Macmillan, 2011).

divides decision-making strategies into System 1 and System 2. System 1 is fast and easy, often relying on emotional responses and gut feelings; System 2 is more effortful and deliberate. Although these strategies do not occur in distinct parts of the brain, both types provide useful constructs to help us understand the kind of processes that are engaged when making distinct kinds of decisions. To be clear, System 1 does not represent an inherently flawed system. On the contrary, it works quite well most of the time; that is why it is so easy, automatic, and effortless to rely on so heavily. When it errs relative to normative theory, however, it tends to do so in predictable and systematic ways that allow insight into the kinds of mistakes people make in undertaking certain kinds of judgments and decisions.

Decision-making in the realm of emerging technology will prove doubly hard because such leaders surround themselves with loyalists rather than competent technocrats.

To preface an argument that will be developed later, these strategies evolved precisely because they serve especially essential functions. For example, utilizing System 1 thinking for most decisions saves costly cognitive energy. Such processes employ feelings that are both genetically instantiated as well as developed through direct experience, including somatic experiences of reward and punishment.⁴ These kinds of feelings do not require complex analytic calculations; they are fast and easy and can be quite powerful, and they provide protection in most situations in the modern world. Such strategies, however, can become problematic when we are confronted with complex decisions that demand, or should require, analysis of data or statistics to reach a proper conclusion; this is where we need System 2 thinking to increase our likelihood of making better choices. Individuals who put as much thought into what to wear as how to fight a predator would likely not have survived to produce many offspring. Therefore, it is easy to see why we

want and expect leaders to engage more deliberate and thoughtful System 2 decision-making schemes when contemplating the risks of nuclear war or AI. Especially under time pressure, however, the automatic urge is likely to do exactly the opposite and privilege feelings over facts.⁵

Leaders who have more personalistic control of their countries are less likely to employ System 2 thinking simply because they do not have to do so to convince the kind of wider governing coalitions characteristic of democratic regimes. These leaders are less institutionally constrained by opposition power or even their domestic populations. Decision-making in the realm of emerging technology will prove doubly hard because such leaders surround themselves with loyalists rather than competent technocrats. Indeed, one of the few genuine weak spots in a personalist's regime in terms of keeping power results from the incompetence (as well as corruption) that characterizes such regimes. Thus, those who might be able to understand the risks posed by emerging technologies may not exist in such regimes, or, if they do, may not be able to exert a decisive influence if they offer something the leader does not want to hear.

The recognition that they might be wrong, or might not know something of great importance, may prove especially challenging for leaders who have been successful enough in their past to achieve and maintain their position in power and thus come to trust their instincts over more abstract data. Personalist leaders, in prizing loyalty over competence, make themselves even more vulnerable to the risks posed by emerging technology not only through their own lack of understanding but also because they will not be surrounded by people who can and will warn them of the inherent risks posed by such innovations. As a result, the ubiquitous nature of the psychological biases discussed below will have a greater sway over personalist leaders than over those subject to more institutional constraint, and this will be particularly so in the arena of emerging technologies, where few people will truly understand the full nature of the innovations. Therefore, much of the risk, benefit, and capacity of these technologies is often unknown prior to wide implementation. Technical mistakes, including a misunderstanding of the limits, or lack thereof, are much more likely with new technologies than with those that have been tried and true across millions of people over many years, such as the radio, telegraph, telephone, or television.

A number of important and influential biases have been documented empirically in psychological

4 Antonio Damasio, *Descartes' Error* (Random House, 2006).

5 Paul Lehner, M-M. Seyed-Soloforouh, Michael F. O'Connor, Stephen Sak, and Theresa Mullin, "Cognitive Biases and Time Stress in Team Decision Making," *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans* 27, no. 5 (1997): 698–703.

research.⁶ The most commonly discussed of the biases in judgment—which are assessments about the probability of outcomes—include representativeness, availability, and the anchoring and adjustment heuristic, each of which has received thorough documentation and generated a great deal of discussion about wider policy implications in previous literature. Likely the most famous and influential aspect of this body of work revolves around prospect theory, the model of decision-making under conditions of risk that was noted in the text of the Nobel Prize awarded to Kahneman. This book is one of the most well-known and widely cited works across all the social sciences.

As significant and meaningful as these biases are, less well-known but equally well-documented biases also deserve attention when it comes to decision-making in the nuclear or emerging technology realms. As presaged previously, these biases include overconfidence, the planning fallacy, and the illusion of validity. These are especially important when considering decision-making around the development, implementation, and use of emerging technologies, where so much remains unknown about potential risks and future innovations. Each of these factors holds enormous influence on the prospects for strategic stability, especially if technologies such as AI are put in full control of assessing and responding to threats and making decisions about the use of nuclear weapons or other weapons of mass destruction such as biological or chemical weapons.

Overconfidence

Overconfidence is one of the overarching biases that tend to infiltrate many other aspects of decision-making, particularly in the area of forecasting.⁷ This effect has been shown to be absolutely endemic across domains, in areas including social prediction,⁸ judgments about frequency and probability,⁹ case studies,¹⁰ financial decision-making,¹¹ and national secu-

arity.¹² Interestingly, overconfidence has been shown to affect the pleasure experienced in performance as well, such that people who are overconfident tend to enjoy their successes less than they would if they had lower expectations for themselves.¹³

Basically, overconfidence means that someone is more confident than they are accurate. A person who thinks he will get a perfect score on an exam and only gets 90 percent is, by definition, overconfident, no matter how good his individual performance is relative to others. Why do some people tend toward overconfidence? First, it is simply easier for people to generate reasons why they are right than to produce alternative hypotheses or disconfirming evidence for why they might be wrong. This tendency also reflects another well-known effect, the confirmation bias, wherein people search for information that validates their prior beliefs while simultaneously discounting evidence that runs counter to those convictions.¹⁴

Second, the foundation of accuracy rests on the ability not only to formulate predictions but also to receive and accept feedback so that future adjustments can be made to improve predictions based on past discrepancies between judgment and outcome. This ability for improvement is why, for example, weather forecasters tend to be adept at short-term precipitation forecasts: They make frequent predictions that receive immediate feedback, so they have many opportunities to learn where and how they err. Despite this success, they have a harder time predicting other important, although rarer, weather events, such as hurricanes or tornadoes, illustrating the complexity of forecasting, even with repetition and immediate feedback. Most forms of social judgment, however, do not allow for both frequency and immediate feedback to occur in close temporal proximity. Indeed, very few social situations occur in similar enough fashion to allow confident conclusions that improve future predic-

-
- 6 Daniel Kahneman, Paul Slovic, and Amos Tversky, eds., *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge University Press, 1982); Amos Tversky and Daniel Kahneman, eds., *Choices, Values, and Frames* (Cambridge University Press, 2000).
- 7 Hal R. Arkes, “Overconfidence in Judgmental Forecasting,” in *Principles of Forecasting*, ed. J. Scott Armstrong (Springer, 2001), 495–515. Note that the generic discussion on overconfidence that follows draws heavily from this discussion.
- 8 David Dunning, Dale W. Griffin, James D. Milojkovic, and Lee Ross. “The Overconfidence Effect in Social Prediction,” *Journal of Personality and Social Psychology* 58, no. 4 (1990): 568.
- 9 Lyle A. Brenner, Derek J. Koehler, Varda Liberman, and Amos Tversky, “Overconfidence in Probability and Frequency Judgments: A Critical Examination,” *Organizational Behavior and Human Decision Processes* 65, no. 3 (1996): 212–19.
- 10 Stuart Oskamp, “Overconfidence in Case-Study Judgments,” *Journal of Consulting Psychology* 29, no. 3 (1965): 261.
- 11 Catherine Schrand and Sarah Zechman, “Executive Overconfidence and the Slippery Slope to Financial Misreporting,” *Journal of Accounting and Economics* 53, nos. 1–2 (2012): 311–29.
- 12 Jeffrey Friedman, “The World Is More Uncertain Than You Think: Assessing and Combating Overconfidence Among 2000 National Security Officials,” *Texas National Security Review* 8, no. 4 (2025): 34–48, <https://doi.org/10.1353/tns.00011>.
- 13 A. Peter McGraw, Barbara A. Mellers, and Ilana Ritov, “The Affective Costs of Overconfidence,” *Journal of Behavioral Decision Making* 17, no. 4 (2004): 281–95.
- 14 Raymond Nickerson, “Confirmation Bias: A Ubiquitous Phenomenon in Many Guises,” *Review of General Psychology* 2, no. 2 (1998): 175–220. For an interesting application to intelligence work, see R. Jervis, “Why Postmortems Fail,” *Proceedings of the National Academy of Sciences* 119, no. 3 (2022): e2116638118.

tions to be drawn. In addition, failed predictions are often considered one-off occurrences; the cost of protecting self-image in this way lies in the lack of learning that such evidence might supply. And in the realms of nuclear brinkmanship, as well as the use of untested, emerging technology for military purposes, there are very few data points on which to render accurate judgments at all.

The consequences of overconfidence are important. First, it leads people to lean toward System 1, rapid, emotionally based decision-making strategies. Second, overconfidence means that people tend to neglect the base rate (also known as prior probability) to their detriment; statistically speaking, Bayes' Law (put forth by eighteenth-century philosopher Thomas Bayes) states that prior probability needs to be incorporated into any judgments about future probability, even in the face of new information, if prospects for accuracy are to be maximized. This tendency to ignore the base rate is found in many judgmental biases and is key to the problems resulting from the better-known representativeness heuristic, whereby people systematically underestimate the probability that an individual belongs to a particular class because they base such judgments on similarity rather than probability.

Overconfidence refers not only to assessments of one's own performance, but also to one's performance relative to others.

Lastly, overconfidence feeds on itself. In leadership circumstances, this tendency toward overconfidence becomes exacerbated where power is asymmetric, as is the case with personalist leaders, especially because such leaders privilege loyalty over competence in their advisers. Sycophants rarely challenge autocratic leaders' judgments, even when those judgments tend toward the absurd, often out of fear for their lives as well as their jobs, depending on circumstance. This tendency means that such leaders rarely have the best information or advice made available to them before making decisions. It also may prove particularly problematic in the face

of emerging technologies because overconfidence flourishes in the face of uncertainty; if someone is not sure how a technology will perform, it is often easier, especially for early adopters and advocates, to champion its many benefits without necessarily considering potential pitfalls, particularly if such recognition might cost profits for those who develop and produce the technology. For example, a number of defense contractors and others in the military were quite overconfident about the performance of the Osprey V-22 aircraft before numerous crashes cost over sixty lives; even though problems with various aspects of the engine and clutch systems proved irrefutable after these accidents, early promoters argued that such challenges could be overcome and were worth the known risks. Subsequent repeated groundings have proved the early proponents incorrect and overconfident in their original estimates.

Overconfidence on the part of leaders may also have some benefits, however. Expressed confidence on the part of leaders encourages fence-sitters to join in a fight against an enemy, and a larger group is then more likely to win without incurring the physical costs of fighting if they can get the other side to back down. This approach could have been an effective combat strategy in our evolutionary past, as well as in more recent conflicts where trying to develop a strong alliance in support of one's cause can prove crucial to survival if not victory, as in the case of the Ukrainian fight against the Russians.¹⁵ And there is no better way to bluff followers and opponents than to deceive oneself, which reduces the risk for behavioral leakage, whereby individuals display nonverbal signals that betray their true underlying beliefs.¹⁶ Indeed, as Steve Rosen notes in *War and Human Nature*, tyrants prove especially susceptible to overconfidence because of the very selection processes that brought them into power; because exaggerated confidence may have worked for them in the past, particularly in gaining power to begin with, such leaders come to trust their instincts and gut feelings as the best source of information, even in the face of evidence to the contrary.¹⁷

Overconfidence refers not only to assessments of one's own performance, but also to one's performance relative to others. Some experimental evidence indicates that people tend to overestimate their performance on difficult tasks, but also think they are worse at that task than others are, while simultaneously underestimating their performance on easy tasks but

15 Richard Wrangham, "Is Military Incompetence Adaptive?," *Evolution and Human Behavior* 20, no. 1 (1999): 3–17; Dominic D. Johnson, *Strategic Instincts: The Adaptive Advantages of Cognitive Biases in International Politics* (Princeton University Press, 2020).

16 Peter Schwarzmann and Joel Van der Weele, "Deception and Self-Deception," *Nature Human Behaviour* 3, no. 10 (2019): 1055–61.

17 Stephen Peter Rosen, *War and Human Nature* (Princeton University Press, 2009).

believing they are better at it than others.¹⁸ Of course, perception of how difficult a task is may lie in the mind of the beholder, but relative assessments can certainly affect contentious decision-making between leaders, particularly if part of what each person wants to achieve is dominance over the others in terms of international power, prestige, and status. Certainly, overconfidence remains endemic, particularly in men. For example, experimental simulated war games show a similarly strong dynamic toward overconfidence among players.¹⁹ Importantly, those who were more overconfident were also more likely to attack their opponents because they were more likely to believe they would win, and this tendency was particularly pronounced among males. A clear analogy to overconfidence in the realm of emerging technology appears obvious: If one side believes their technology is superior, rightly or wrongly, they are more likely to believe in the value and utility of that technology than if they believe they remain at a technological disadvantage. When both sides feel that they are superior, the risk of conflict is raised because each side believes they can best the other with their advanced technology, but inevitably one side will be wrong in their assessment.

Several proposals have been put forward that might help mitigate the effects of overconfidence on judgment.²⁰ Some of these suggestions are reminiscent of best practices in intelligence gathering and analysis.²¹ These recommendations include encouraging people to consider alternative hypotheses and explanations for events or behavior they may not understand, particularly if confronting a novel situation or circumstance, which the contemplation of nuclear war or of AI taking over all human decision-making tasks certainly constitutes. Second, individuals should explicitly consider why they might be wrong. This technique often inspires people to see aspects of a situation they might not have seen otherwise, including their opponent's viewpoint; this may, however, prove exceedingly difficult for some leaders. Third, having a devil's advocate in any decision-making group can help everyone contemplate alternative or outside perspectives, although all too often the devil's advocate can be captured by the beliefs of the group for social reasons if they face ostracism for posing disagreement. Devil's advocates can also be

easily marginalized in many groups or situations. For example, when President Jimmy Carter was making his decision about the rescue mission of the hostages in Iran in 1980, he became increasingly frustrated by then-Secretary of State Cyrus Vance's opposition. In the end, Carter simply held decisive meetings at times he knew Vance was out of town or otherwise unable to attend.²² Yet another helpful suggestion is to develop a strategy that mimics weather forecasters: Make a prediction and then solicit feedback, take the feedback seriously, and if possible, test the prediction in simulated, experimental, or minimal risk contexts before rendering a new decision.

Of course, such strategies all take time, which leaders often do not feel they have, especially in a crisis, even if it is better to be right than to be quick. However, practicing such strategies prior to crisis, and putting in place explicit rules or standard operating procedures that encourage or require such standards of behavior, may reduce the likelihood that overconfidence generates the kinds of faulty judgments, inaccurate predictions, and dangerous outcomes that too often result in its wake, including financial meltdowns and inadvertent and dangerous conflicts. Autocratic leaders are the ones most in need of such strategies because they privilege loyalty over competence. Ironically, however, personalist leaders are those least likely to devote time to such tasks, precisely because such leaders do not feel that they need to do so, not only because of the tendency toward overconfidence, but also because they have fewer institutional constraints.

Planning Fallacy

Daniel Kahneman argues that one of the most pervasive and significant cognitive biases revolves around optimism, which can obviously be seen to operate hand in glove with overconfidence. In many circumstances, optimism can be a quite valuable trait; optimists, for example, live much longer than their glass-half-full counterparts. Using one's imagination to achieve a positive sense of mastery or control enhances one's sense of well-being.²³ Optimism can also lead to an illusion of control,²⁴ however, whereby individuals feel that they have a lot more control over the outcome of events than

18 Don A. Moore and Paul J. Healy, "The Trouble with Overconfidence," *Psychological Review* 115, no. 2 (2008): 502.

19 Dominic D. P. Johnson, Rose McDermott, Emily S. Barrett, Jonathan Cowden, Richard Wrangham, Matthew H. McIntyre, and Stephen Peter Rosen, "Overconfidence in Wargames: Experimental Evidence on Expectations, Aggression, Gender and Testosterone," *Proceedings of the Royal Society B: Biological Sciences* 273, no. 1600 (2006): 2513–20.

20 Friedman, "The World Is More Uncertain Than You Think"; Arkes, "Overconfidence in Judgmental Forecasting," 495–515.

21 Robert Jervis, *Why Intelligence Fails: Lessons from the Iranian Revolution and the Iraq War* (Cornell University Press, 2010).

22 Rose McDermott, *Risk-Taking in International Politics: Prospect Theory in American Foreign Policy* (University of Michigan Press, 2001).

23 Shelly E. Taylor and Jonathon D. Brown, "Illusion and Well-Being: A Social Psychological Perspective on Mental Health," *Psychological Bulletin* 103, no. 2 (1988): 193.

24 Seyom Brown, *The Illusion of Control: Force and Foreign Policy in the 21st Century* (Rowman & Littlefield, 2004).

they actually do. Kahneman argues that this bias provides welcome and necessary ballast against the even more powerful force of loss aversion, whereby people weigh losses more heavily than commensurate gains psychologically. If people did not have an ability to remain optimistic in the face of incipient threat, they might become too paralyzed to act, thus ensuring destruction. Too much optimism, however, can also lead to the neglect of crucial features of a problem.

Interestingly, some experimental evidence suggests that, although people underestimate the time it will take them to complete tasks, they do not make this mistake when predicting completion time for others.

Indeed, one important manifestation of this bias toward optimism lies in the planning fallacy, which has most often been applied to the time people predict it will take to complete tasks, although this tendency holds much wider implications.²⁵ Take a simple example: Most academics woefully underestimate the time it will take to finish a particular paper or project. Because planning and prediction are so closely intertwined, the planning fallacy can lead to overly optimistic forecasts and expectations of success.²⁶ As a result, this tendency can push individuals to exaggerate potential benefits while simultaneously dismissing expected costs.²⁷ This outcome can encourage people to take on needlessly foolish, risky, or costly projects, including entering wars, in part because they expect things to take less time and be less costly and more successful than an accurate forecast might suggest. The Russian invasion of Ukraine provides a grim recent example of this phenomenon; Putin clearly not only proved

overconfident but also failed in his planning and prediction of success. He expected the war, which is about to enter its fourth year, to be over in three days. He, like many others including the American administration, vastly underestimated Ukrainian resolve and willingness to fight while simultaneously overestimating Russian strategic planning and technological superiority. But a plan for an easy three-day victory does not easily survive a three-plus-year war of grinding attrition.

Interestingly, some experimental evidence suggests that, although people underestimate the time it will take them to complete tasks, they do not make this mistake when predicting completion time for others. This tendency appears to result from the fact that, when making assessments about themselves, people focus on their future plans, rather than their past experience with similar tasks, which would serve as a better predictor.²⁸ Once again, failure to incorporate the base rate into prediction leads to predictable, systematic bias in judgment. In his contribution to this Roundtable, Michael Horowitz discusses the tendency to overestimate the value of emerging technology, particularly in the short term. A recent salient example derives from the acclaim that greeted the first deployment of widespread generative AI with OpenAI's ChatGPT. While the initial assessments promised both the solution to all human problems as well as the massive loss of jobs, each new iteration has not been met with similarly near-universal acclaim, particularly because none have offered a qualitative leap beyond the first iteration as expected. The most recent rollout of version 5 of ChatGPT has garnered relief in some camps that we have not quite entered the matrix yet, while others express disappointment that advances have not been as rapid as the initial launch suggested.

The attributional discrepancy between what we believe we can accomplish and what we expect others to be able to do is important because it highlights one of the common themes in the vast majority of judgmental biases: that people do not tend to apply them equally to themselves and to others. This asymmetry likely represents an aspect of the fundamental attribution error, whereby individuals overestimate the extent to which they are subject to situational pressures while simultaneously underestimating

-
- 25 Roger Buehler, Dale Griffin, and Johanna Peetz, "The Planning Fallacy: Cognitive, Motivational, and Social Origins," in *Advances in Experimental Social Psychology*, ed. Mark Zanna and James Olson, vol. 43 (Academic Press, 2010), 1–62.
- 26 Roger Buehler and Dale Griffin, "The Planning Fallacy: When Plans Lead to Optimistic Forecasts," in *The Psychology of Planning in Organizations*, ed. Michael Mumford and Michael Frese (Routledge, 2015), 47–73.
- 27 Roger Buehler, Dale Griffin, and Michael Ross, "Inside the Planning Fallacy: The Causes and Consequences of Optimistic Time Predictions," in *Heuristics and Biases: The Psychology of Intuitive Judgment*, ed. Daniel Kahneman and Thomas Gilovitch (Cambridge University Press, 2012), 250–70.
- 28 Roger Buehler, Dale Griffin, and Michael Ross, "Exploring the Planning Fallacy: Why People Underestimate Their Task Completion Times," *Journal of Personality and Social Psychology* 67, no. 3 (1994): 366.

the extent to which others are, as well as assuming they are less (and others are more) influenced by their character or personality. This bias encourages leaders to infer a level of intentionality, often hostile, to others' behavior where it may not in fact be warranted.²⁹ The typical person believes "I did what I did" because the circumstances demanded it, but "you did what you did" because that is who you are as a person. This tendency gives me a reasonable excuse for bad behavior while allowing me to blame you for doing terrible things because I believe that is who you are as a person. Any failed weapons design illustrates this phenomenon, whereby some early supporters strive to achieve a qualitative improvement in design or function, but failure demonstrates that the planning process was flawed, often because of overconfidence among engineers and other developers who misunderstood the difficulty of the challenges they confronted.

A couple of strategies might help diminish the effect of the planning fallacy on prediction. One strategy asks people to explicitly state their intentions for implementing their plans prior to undertaking the task. In other words, they have to say what they plan to do, why they plan to do it, and what effect they believe the action will have on the outcome. This approach tends to reduce unwarranted optimism.³⁰ The second strategy encourages people to "unpack" the component elements involved in larger tasks to help them understand that more may be involved in their planning than they realize and more gaps may need to be filled prior to implementation.³¹ The challenge, in today's nuclear environment, is that the willingness and institutional capacity to adopt these types of fixes vary widely across regimes and individual leaders, and thus some states may be more willing to adopt these strategies than others.

The Illusion of Validity

The illusion of validity refers to a bias wherein people overestimate their ability to predict an outcome from data, particularly when that data appears to tell a coherent narrative story.³² This tendency is related to the representativeness heuristic because both relate to how people judge outputs based on their similarity to inputs. In this case, people's predictions are based more on the "fit" in a narrative

between input and output than on the accuracy of the relationship between them. Amazingly, this effect persists even when people are made aware of the ways in which the accuracy of their judgments might be compromised. The classic example given by Amos Tversky and Daniel Kahneman involves job interviews, where observers remain highly confident in their assessments of given candidates they have met, despite knowing the reams of evidence showing that such personal assessments are highly flawed and inaccurate. The ubiquity of such interviews proves that even educated people dismiss data in favor of their own, however biased, judgment.

Although certainly some of this bias rests on a simple desire to assess likability, it also betrays both overconfidence in personal judgment and the insidious nature of these biases, even among the well educated and well informed. Good examples of this phenomenon include Trump's assessments of world leaders such as Putin and North Korean President Kim Jong Un based on personal encounters, privileging such experiences over decades of America's prior negative interactions and experiences with the two countries. It is easy to see how some leaders overvalue personal experiences of flattery or humiliation more than substantive political, strategic, or military concerns. In short, elites are not impervious to the more negative manifestations of these biases. For example, Trump claimed that he was "not worried" about the display of Chinese military exercises around Taiwan in late 2025, arguing that these drills represented routine activities. He went on to state that the reason he was not concerned was because he "had a great relationship with President Xi."³³ This privileging of personal relationships over objective assessments offers a classic illustration of this phenomenon.

Importantly, consistency in a perceived pattern leads people to be particularly prone to this bias. In this way, individuals easily favor a good or consistent story over more accurate information that does not make sense or hang together in a meaningful narrative, often because information is missing and it requires more cognitive effort to try to put the pieces together. A great illustration of this bias comes from a study by Scott Plous that showed how individuals rated the likelihood of nuclear war to be much higher when the story about what precipitated it seemed

29 Lee Ross, "From the Fundamental Attribution Error to the Truly Fundamental Attribution Error and Beyond: My Research Journey," *Perspectives on Psychological Science* 13, no. 6 (2018): 750–69.

30 Sander Koole and Mascha van't Spijker, "Overcoming the Planning Fallacy Through Willpower: Effects of Implementation Intentions on Actual and Predicted Task-Completion Times," *European Journal of Social Psychology* 30, no. 6 (2000): 873–88.

31 Justin Kruger and Matt Evans, "If You Don't Want to Be Late, Enumerate: Unpacking Reduces the Planning Fallacy," *Journal of Experimental Social Psychology* 40, no. 5 (2004): 586–98.

32 Amos Tversky and Daniel Kahneman, "Judgment Under Uncertainty: Heuristics and Biases," *Science* 185, no. 4157 (1974): 1124–31.

33 Agence France Presse, "Trump Says Not Concerned by Chinese Drills Around Taiwan," *Barron's*, December 29, 2025, <https://www.barrons.com/news/trump-says-not-concerned-by-chinese-drills-around-taiwan-c8ce7ee4>.



logically cohesive, even when it was more specific and thus less likely statistically than a more generic judgment.³⁴ Specifically, people judged a nuclear war breaking out after a conflict between the then–Soviet Union and NATO to be more likely than a nuclear war breaking out in general. Of course, since the former is merely a subcategory of the latter, the latter is more likely overall but does not seem as likely because it lacks a causal story. Although this study was presented as an example of anchoring—wherein people tend to place more weight on an initial assessment even if they know it to be irrelevant to the current judgment—it also illustrates how powerful narratives can be in influencing judgment and choice. This correlation implies, then, that early advocates of a new technology will be much more able to convince others of its utility if they can tell a clear and compelling, ideally emotionally evocative story about why it will be valuable, perhaps by claiming that it will save lives or reduce costs. Other potential negatives might then be dismissed. So, for example, the enormous energy and water drains caused by AI demands receive much less attention because its promoters claim that AI will make everyone’s lives easier and more productive, never mind that such

wonderful lives will be filled with droughts, power outages, and higher energy costs.

Skillful politicians can be exceptionally good at weaving a convincing story about the evil intent of an enemy or why a particular belligerent response is justified, even when these assessments are not accurate. Leaders may argue that an adversary’s adoption of an innovative technology is more threatening than it is, simply because they want an excuse to justify aggressive actions that are desired for other reasons. In addition, as we know from the Russian interferences in the 2020 American presidential election, it is possible that the disruptive nature of recent technologies, such as chatbots, are not fully understood during their initial deployment. Accompanied by emotional content and powerful visual images, such narratives of threat can prove especially convincing to gullible, uneducated, or emotionally aroused publics.

The Prominence Effect

The final phenomenon to be considered here, known as the prominence effect, may also lead to biased decisions pertaining to the use of nuclear weapons or other emerging technologies, and therefore to a departure

34 Scott Plous, “Thinking the Unthinkable: The Effects of Anchoring on Likelihood Estimates of Nuclear War,” *Journal of Applied Social Psychology* 19, no. 1 (1989): 67–91.

from rationality in ways that challenge strategic stability. This effect results from the need to justify one's actions to oneself as well as others.³⁵ When a leader must defend their decisions, such a discussion focuses attention on the use of aggression in terms of the defense and protection it provides to their own citizens. Thus, enormous amounts of violence can be readily justified if that violence is understood and presented as serving the overriding imperative of population protection and national security.

The desire for vengeance in the face of an attack may prove strong and immediate, and retaliation might be launched before the full consequences are considered.

The prominence effect highlights another important characteristic of decision-making. People seem to have a clear and intuitive sense of hierarchy in their values, and these assessments prioritize personal safety and security over other values, including money, no matter how excessive or disproportional the harm to others that might result from the efforts taken to obtain it. However, in a world of intercontinental ballistic missiles, second-strike capability, and increasingly sophisticated artificial and cyber technologies, the extent to which punishing others may redound quickly to self-annihilation seems lost. The desire for vengeance in the face of an attack may prove strong and immediate, and retaliation might be launched before the full consequences are considered. Particularly as advances in technology increase the time pressure on leaders, catastrophic actions might be taken without full recognition of the way the use of weapons of mass destruction risks the sixth, so-called Anthropocene extinction, of humans.

Better and more extensive education about the extent and capacity of retaliatory damage may help diminish support for the use of nuclear weapons, at least in part. Indeed, the consequences of a nuclear war are so catastrophic that they stretch far beyond

the ability of humans to comprehend, and the risks that attend to the adoption of new technologies may similarly challenge human understanding. This threat is not simply limited to nuclear weapons; massive cyberattacks that take out the water or power supplies of large populations could produce mass casualty events. More significantly, gain-of-function experiments in biological weapons would make COVID-19 look like child's play and could easily lead to the extinction of the human race, along with other species. The risk of AI for human existence has not yet begun to be seriously considered, but—in conjunction with automaticity in the use of nuclear or other weapons of mass destruction—it could threaten wide swaths of humanity if not human existence itself. Even to a much smaller degree, we are seeing examples of ChatGPT and other AI “helping” vulnerable individuals to kill themselves and others more effectively; several notable lawsuits have already been filed against OpenAI by distressed relatives. Moreover, the head of medicine at University of California San Francisco recently said that as good as ChatGPT is on a lot of medical advice, in several examples of answers he considered he judged that following such advice would risk the life of the questioner. The vastness of these effects precludes any possibility for a rational cost-benefit analysis of risks and benefits; indeed, the only “rational” choice is not to use many of these technologies, but of course that is unrealistic in the modern world.³⁶

The power of the prominence effect also shows that leaders can easily manipulate susceptible publics into supporting aggressive enterprises by cloaking them in the veil of national security, a concept that is already both vague and impervious to public transparency and questioning. Specifically, when people assess the degree of risk of an enterprise, those risks that are unknown and those that are considered “dread” risks are judged as much riskier than other forms. So, for example, car crashes kill a large number of people every year, but automobiles are a well-known technology, so they are judged to be less risky than a technology that is less well understood, like cyber or AI.³⁷ Thus, it is easier for leaders to induce fear in followers by framing threats in terms of such unknown or dread risks. In this way, leaders can easily reframe a personalistic fight they undertake for their own regime security as a national security requirement, with many none the

35 Paul Slovic, “When (In)Action Speaks Louder Than Words: Confronting the Collapse of Humanitarian Values in Foreign Policy Decisions,” *University of Illinois Law Review Slip Opinions* (April 2015): 24–31.

36 Paul Slovic and Herb Lin, “The Caveman and the Bomb in the Digital Age,” in *Effects of the Global Information Ecosystem on the Risk of Nuclear Conflict*, ed. Herbert Lin, Benjamin Loehrke, and Harold Trinkunas, vol. 7 (2018), from a workshop conducted at the Center for International Security and Arms Control, Stanford University.

37 Paul Slovic, Baruch Fischhoff, and Sarah Lichtenstein, “Perceived Risk: Psychological Factors and Social Implications,” *Proceedings of the Royal Society of London: A. Mathematical and Physical Sciences* 376, no. 1764 (1981): 17–34.

wiser. This scenario poses a particularly salient and dangerous specter in which leaders can draw publics into supporting so-called “preventive” attacks by claiming a preemptive need for self-defense, much as the Bush administration did in justifying the war in Iraq in 2003 under the pretext of eliminating weapons of mass destruction, or as Putin did in justifying the invasion of Ukraine in 2022.

The Role of Evolution

The work of Kahneman and Tversky represented the apex of the cognitive revolution in psychology, and their meticulous and elegant experimental demonstration and documentation of heuristics and biases revealed the important ways in which humans make systematic errors in decision-making in predictable ways under specific circumstances.³⁸ Later evolutionary approaches in psychology took issue with the notion of the human mind as error prone and mistake riddled, noting that humans are not so much irrational as they are adaptive to given circumstances.³⁹ Thus, humans are more ecologically rational than economically rational. These perspectives have largely been at intellectual and theoretical loggerheads, so it is important to note that in the domain of combat, war, violence, and conflict, they need not be.

Specifically, the psychological mechanisms that have allowed humans to operate effectively and efficiently under resource constraints of time and energy required certain trade-offs between accuracy, speed, and energy. Natural selection operates to optimize prospects for survival, and this has resulted in some universal and predictable biases. This development does not mean that every error demonstrated experimentally derived from evolutionary pressures, nor that every evolutionary strategy operates in optimal fashion all the time. Rather, evolution works off large-scale odds averaging across enormous numbers of people over long periods of time; human psychological architecture develops around those strategies that have worked best for most people most of the time under certain circumstances in the past to increase the odds of genetic survival. Even tiny advantages in predispositions can be preserved under such a system. Of course, this strategy does not mean that features that

worked well in the past will necessarily work well in the future. However, remaining blind to these features of human psychology does not make them go away, nor does it increase the odds of survival going forward. In other words, strategies for revenge that may have worked in the past to preserve populations from enemies may now doom all of humanity to destruction with the powerful advances that have been made in weaponry; the human mind does not evolve as rapidly as technology, and too often the embrace of such technology occurs far earlier than consideration of its potentially destructive effects.

There are many circumstances in which what appear to be malfunctioning biases can actually improve decision-making.⁴⁰ In the simplest of examples, invoking System 1’s fast, emotion-based processing saves a lot of cognitive energy that would otherwise be unnecessarily spent on mundane tasks such as deciding what to eat or what to wear. But more significantly, in a world of uncertainty and inequality, evolved cognitive biases can sometimes, although not always, improve decision-making in certain realms. Many aspects of human psychological architecture that appear to be flaws, or bugs, are instead more accurately characterized as design features, helping solve ancestral rather than modern problems to maximize prospects for fitness.⁴¹ Notably, however, there can be clashes between the requirements and demands of the ancestral past and the challenging circumstances of the modern environment. The most ubiquitous example derives from a preference for sugar: Ancestors who possessed this trait in a world of food scarcity, for example by preferring honey to greens, were much more likely to survive and reproduce, whereas in the modern world, defined by couches and fast food, such preferences lead to obesity, diabetes, and other health concerns.

Similarly, historical trade-offs between speed, accuracy, and energy that worked well in solving ancestral problems may not operate in the same way in the modern context. In the past, preserving energy was key to survival in a world of resource scarcity. Now people need to spend more energy to preserve health, although most of us have inherited a preferred tendency toward laziness, so expecting energetically costly actions is not necessarily realistic. But the

38 Amos Tversky and Daniel Kahneman, “Judgment Under Uncertainty: Heuristics and Biases; Biases in Judgments Reveal Some Heuristics of Thinking Under Uncertainty,” *Science* 185, no. 4157 (1974): 1124–31; Daniel Kahneman and Amos Tversky, “Prospect Theory: An Analysis of Decision Under Risk,” in *Handbook of the Fundamentals of Financial Decision Making: Part I*, ed. Leonard C. MacLean and William T. Ziemba (World Scientific Publishing, 2013), 99–127.

39 Jerome H. Barkow, Leda Cosmides, and John Tooby, eds., *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (Oxford University Press, 1995).

40 Dominic D. P. Johnson, Daniel T. Blumstein, James H. Fowler, and Martie G. Haselton, “The Evolution of Error: Error Management, Cognitive Constraints, and Adaptive Decision-Making Biases,” *Trends in Ecology & Evolution* 28, no. 8 (2013): 474–81.

41 Martie G. Haselton, Daniel Nettle, and Damian R. Murray, “The Evolution of Cognitive Bias,” *The Handbook of Evolutionary Psychology* (2015): 1–20.

problem now is that speed operates at an entirely different pace than in the past. Privileging accuracy in the modern world simply falls by the wayside as result of increasing time pressures imposed by emerging technologies. In the past, accuracy would never have been as important for survival as maintaining one's social network; believing falsehoods not related to food and fidelity would have imposed few costs. Therefore, that impetus is the first to go as time pressure increases and humans remain fundamentally committed to energy conservation.

Policy Implications

Just as the sugar preference example above illustrates, the psychological biases inherent to human psychological architecture result in meaningful mismatches with the demands of the modern world in other ways as well, including the nature of personalist leadership and the role of emerging technologies. First, authoritarian leadership structures that might work for the benefit of all in a group of 150 or so—where everyone is aware of what the leader is doing and coalitions of men can kill a leader who does not work for the benefit of all—are not viable in an environment of millions of constituents who not only cannot keep track of everything the leader is doing, but also will not have the collective ability to overthrow corrupt leaders easily or quickly.⁴² Second, new technologies in the past have sometimes proved beneficial, as in the case of electricity or indoor plumbing, for example, but some have also led to more effective and efficient ways of killing enemies, from the Chinese development of gunpowder to nuclear weapons. Even lethal weapons that were geographically limited in scope had less destructive potential than modern ones with global reach. Even past pandemics were naturally occurring and not produced in a lab. Therefore, the scope of destructive potential posed by such weapons is also a mismatch for the more localized conflict that characterized the majority of human evolutionary history.

Even if the psychological biases that served humans well in the past no longer serve the needs of the current environment, we cannot simply evict them from human nature. Even if we wanted to do so, it would not be wise, because too much of what works well about human nature would disappear as well. Rather, we need to take steps to ensure that these biases do not lead us into catastrophic decisions that can harm enormous numbers of people. In each subsection above, strategies to help overcome the particular bias mentioned have been outlined. The goal here has been to alert decision-makers to the need to develop measures, procedures, and protocols to help prevent, or at

least reduce the likelihood of, these costly biases from influencing decisions in unnecessarily negative ways. Unfortunately, these costly biases are especially likely to appear in new or unfamiliar domains, including the use of emerging technology, and they are least likely to be mitigated by some of the regimes and personalist leaders that now possess nuclear weapons—an especially combustible combination of risks.

Even if the psychological biases that served humans well in the past no longer serve the needs of the current environment, we cannot simply evict them from human nature.

There is, however, a more overarching and comprehensive insight from the work on decision-making biases that should also be mentioned and advocated. Left to their own devices, most humans will rely on System 1 decision-making most of the time to the extent possible. Because System 1 is fast and easy, it requires much less thought and effort and energy. As noted earlier, this system works well most of the time, but can cause systematic and predictable problems when an analysis of abstract data, statistics, complex decisions, or causal forces that are distant in time and space is required. System 1 thinking can also encourage biases, such as overconfidence.

Engaging System 2 thinking will slow things down and make much more deliberate and thoughtful analysis and decision-making possible. But getting people to use System 2 is neither easy nor automatic and indeed is least likely when it is most necessary—under conditions of genuine threat, uncertainty, and time constraint. In addition, System 2 thinking does not exist without its own biases; it is simply that those biases are much less likely to produce problems under the conditions of the time-constrained decision-making that is most likely to characterize the lead-up to nuclear use. Increased time pressure will increase reliance on System 1 decision-making. Turning such decisions over to AI, however, even if it is faster and more predictable, will not improve matters, because, as noted above, AI simply replicates human errors at scale. AI also has no conscience, so it is not capable of empathy and cannot truly consider the cost to real human beings; it simply exists as a

42 Richard Wrangham, *The Goodness Paradox: The Strange Relationship Between Virtue and Violence in Human Evolution* (Vintage, 2019).

manifestation of past human decisions, which, as we all know, did not often turn out very well in the realm of conflict and war.

As a result, one of the things that might help decision-making under such conditions would be to have established protocols or standard operating procedures in place—not for action but for process—including putting checklists in place that can be invoked in times of crisis.⁴³ As with other protocols, leaders and advisers could practice and train using decision strategies developed in concert with decision analysts to produce the most accurate and least destructive choices possible. Indeed, as a training tool, AI might prove quite useful in undertaking simulations for various scenarios; such training has been shown to be quite valuable. When such checklists were instituted in hospitals to reduce the number of deaths, this measure saved over 100,000 lives a year, even though many of the items appeared completely obvious, such as “make sure to wash hands.”⁴⁴ Developing, testing, and training with sophisticated and flexible decision processes, including those supported by AI, requiring leaders to answer certain questions (such as “What are your critical values?”) prior to authorization to launch nuclear weapons would allow leaders and advisers the opportunity to ensure, even under conditions of enormous stress and time pressure, that they are considering the proper elements of the problem, asking the right questions, seeking the most relevant evidence, entertaining alternative hypotheses, and testing competing predictions based on alternative choices. Making this process more familiar and accessible may not make it automatic, but it should render it less stressful and effortful, and perhaps less politically risky for personalist leaders because they are not turning over decision-making to other individuals but rather gaining support for their own actions. These measures could produce outcomes less prone to bias, as they have in other settings.

Conclusion

Human psychological architecture evolved under near-constant direct experience with conflict, aggression, combat, threats, and risks. Humans possess a series of biases in decision-making under conditions of risk and threat that may have benefited people in the past and may even continue to do so in the present around many issues. In the current nuclear world, however, enabled by an array of rapidly evolving emerging technologies, a single

failure risks the future of humanity itself, and such biases may no longer serve us well to generate the best outcomes, especially under conditions of stress, conflict, uncertainty, and time pressure.

Invoking the slower, more thoughtful, and sophisticated decision-making mechanisms available is therefore important for producing a more deliberate way to improve our prospects for successful decision-making in the realm of nuclear (and conventional) conflict. Developing comprehensive standards for decision-making when structuring national security policy is imperative to ensure that optimal decision-making strategies are employed when they are needed most, especially given the heightened stakes posed by nuclear weapons, emerging technologies, and other weapons of mass destruction, including biological weapons.⁴⁵ ●

Rose McDermott is the David and Mariana Fisher University Professor of International Relations at Brown University and a fellow in the American Academy of Arts and Sciences. She works in the area of political psychology. She received her PhD (political science) and MA (experimental social psychology) from Stanford University and has also taught at Cornell and University of California Santa Barbara. She has held fellowships at the Radcliffe Institute for Advanced Study, the Olin Institute for Strategic Studies, and the Women and Public Policy Program, all at Harvard University, and has been a fellow at the Stanford Center for Advanced Studies in the Behavioral Sciences twice. She is the author of six books, coeditor of two additional volumes, and author of over two hundred academic articles across a wide range of topics, including American foreign and defense policy, experimentation, national security intelligence, gender, social identity, emotion, and biological and genetic bases of political behavior.

Brown University, Providence, RI, USA, email: rose_mcdermott@brown.edu.

Acknowledgments: I would like to thank Paul Slovic for comments on an earlier version of this paper, and Sheena Greitens and Joshua Rovner for comments on a later version. I would also like to thank Harold Trinkunas for his intrepid support of the larger Roundtable from the beginning.

Image: “Atomic Landscape” Nagasaki, Japan, 1946, by Robert Graham.⁴⁶

43 Friedman, “The World Is More Uncertain Than You Think.”

44 Joseph McCannon, Marie Schall, David Calkins, and Alexander Nazem, “Saving 100000 Lives in US Hospitals,” *British Medical Journal* 332, no. 7553 (June 3, 2006): 1328–30.

45 Atul Gawande, *The Checklist Manifesto* (Picador, 2010).

46 For image, see <https://catalog.archives.gov/id/6446321>.