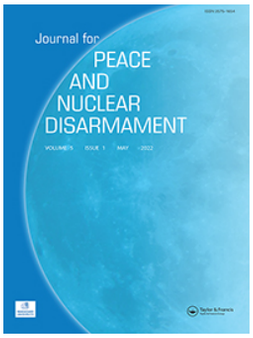

Routledge
Taylor & Francis Group

Journal for Peace and Nuclear Disarmament

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rpnd20

Counterfactual thinking & nuclear risk in the digital age: The role of uncertainty, complexity, chance, and human psychology

James Johnson

To cite this article: James Johnson (2022): Counterfactual thinking & nuclear risk in the digital age: The role of uncertainty, complexity, chance, and human psychology, Journal for Peace and Nuclear Disarmament, DOI: 10.1080/25751654.2022.2102286

To link to this article: https://doi.org/10.1080/25751654.2022.2102286

Published online: 25 Jul 2022.

JOURNAL FOR PEACE AND NUCLEAR DISARMAMENT
https://doi.org/10.1080/25751654.2022.2102286

Counterfactual thinking & nuclear risk in the digital age: The role of uncertainty, complexity, chance, and human psychology

James Johnson

Department of Politics and International Relations, University of Aberdeen, King's College Aberdeen, Aberdeen, UK

ABSTRACT

Will emerging technology increase the possibility of nuclear war? Given the multitude of ways emerging technology intersects with nuclear weapons, critical thinking about an imagined future that goes beyond net assessment, myopic mirror-imaging, and extrapolation of present trends should be a core task of policymakers. This article builds on the notion of "future counterfactuals" to construct imaginative yet realistic scenarios to consider the future possibility of a nuclear exchange. It highlights the critical role counterfactual scenarios can play in challenging conventional wisdom about nuclear weapons, risk analysis, war-fighting, and linear thinking. In emphasizing the role of uncertainty, cognitive bias, and fundamental uncertainty in world politics, the article also contributes to the literature about the risk of inadvertent and accidental nuclear war.

ARTICLE HISTORY
Received 27 March 2022
Accepted 13 July 2022

KEYWORDS
Disruptive emerging technology; nuclear war; counterfactuals; complexity; deterrence theory

Introduction

Will emerging technology increase the risk of nuclear war? Because of the multitude of ways disruptive emerging technology (DET) (Talmage 2019; Johnson 2020; Acton 2015; Sherwood-Randall 2020) intersects with nuclear weapons – and the broader "nuclear deterrence architecture"[1] (Boulanin et al. 2020, 24) – critical thinking about an imagined future that goes beyond net assessment, myopic mirror-imaging, and extrapolation of present trends should be a core task of policymakers (Lebow and Pelopidas Forthcoming).[2] For most defense planners and policymakers, the prospect of a nuclear war has become a political-strategic exercise in denial and complacency, driven by an inflated sense of confidence in the ability to control nuclear escalation and nuclear

CONTACT James Johnson james.johnson@abdn.ac.uk Department of Politics and International Relations, University of Aberdeen, King's College Aberdeen, UK

[1]The "nuclear deterrence architecture," includes early-warning and intelligence, surveillance, and reconnaissance (ISR) systems, command and control, precision strike and delivery, and non-nuclear operations such as cyber, electronic warfare, counter-space, missile defense, and physical security.

[2]A new wave of transformative disruptive emerging technology (DET) – associated the broader "Fourth Industrial Revolution" (4IR) – including artificial intelligence (AI), quantum technology, nanotechnology, hypersonic weapons, additive manufacturing, and directed energy weapons is expected to have a wide-ranging impact at a societal, economic, ethical, and domestic-political level. The implications of dual-use – civilian and military uses – DET's, some of which have not yet reached maturity, is already affecting the stability of traditional deterrence relationships between the West and Russia and between the West and China.

weapons, and understating the role in luck in preventing nuclear detonation – often treating luck, similar to uncertainty, as unquantifiable and thus negligible. In short, there are several practical inconsistencies and epistemic inconsistencies in the current intellectual milieu, which is creating poor intellectual habits that will have real-world, and potentially destabilizing consequences in present and future strategic planning. This attitude has meant that efforts to prevent nuclear war are frequently reduced to abstract technical-capabilities solutions – for example, the modernization of nuclear forces and the pursuit of advanced strategic non-nuclear weapons associated with the "Third Nuclear Age" (Futter and Zala 2021) – in lock-step with the continued affirmation of the logic of rational-based deterrence and – in the case of the United States – the goal of nuclear superiority to guide war planning and disarmament and arms control proposals (Lieber 2008). Put differently; these new whiz-bang technologies do not exist in a vacuum. Instead, to avoid obfuscation, the impact of DET's on nuclear risk needs to be considered in the broader geopolitical, domestic-political, ethical, and psychological context into which they have been supplanted (Horowitz et al. 2018; Kroenig 2021).

The dichotomy between the nuclear deniers and proponents of rational deterrence can in part be explained, *inter alia*, by a) the unprecedented nature of the ultimate weapon (Amis 1987), b) limited empirical data to inform strategic planners and policymakers, c) ideology, vested interests, and "self-censorship",(Pelopidas 2016), d) the psychological preferences of policymakers (Zwald 2013), and e) the existential fear, taboos, and popular cultural associations with the possibility of existential thermonuclear war (Tannenwald 2018; Jacobs 2010). An additional variance in debates about the use of nuclear weapons is those who implicitly view some limited use of nuclear weapons as politically acceptable. In contrast, others view nuclear detonation as unacceptable and thus the notion of "nuclear strategy" as oxymoronic (James 1977). Nuclear studies – like security and strategic studies more broadly – are, therefore, imbued with analytical, normative, and epistemological assumptions and inconsistencies that scholars often overlook. As William James mused: "Concepts, first employed to make things intelligible, are clung to often when they make them unintelligible" (James 1977, 560).

Exposing and challenging these assumptions and the prevailing wisdom about the control and manageability of nuclear risk is an area ripe for the engagement with imaginative "counterfactual" analysis". (Tetlock and Belkin 1996, 4 and 23). According to Richard Lebow, counterfactuals "allow for the construction of rational templates that are used to assess the behavior of real-world actors" (Lebow 2015, 403). Because counterfactuals can illuminate otherwise epistemically and empirically inaccessible (both temporally and spatially) perspectives, they are an essential tool for scholars to formulate theories and test hypotheses.

Empirically and theoretically rigorous counterfactuals can be used, for instance, to re-examine the history of nuclear "close-calls" (Lewis et al. 2014) and accidents to expose the "illusion of control" (i.e. overconfidence in the ability to direct events) (Trivers 2014)[3] perceiving the past over-deterministically, and the focus of this article, the possibility of future nuclear use. While future scenarios are a critical feature of defense planning and

[3]The "illusion of control" is also supported by anthropologic research. Namely, human males have a strong tendency toward overconfidence and illusions of control, believing they have more control than they do. This tendency also provides an evolutionary advantage; those who can bring more people to their side of a fight are more likely to win.

the basis for arms control, disarmament, and broader non-proliferation efforts, they rarely go beyond mere extrapolation of current capabilities and technological determinism (Pelopidas 2020; McDonald and Bell 2019; Connelly et al. 2012). Scholars need, therefore, to use counterfactual scenarios in a more imaginative way that engages the policymaking community to influence perceptions and perhaps behavior. Like Herman Kahn's objective in writing *On Thermonuclear War* in 1960, the goal of counterfactual thinking is to build on the consensus that exists on "what we are trying to avoid" and buy more time (Kahn 1960). This article contributes to the, albeit limited, literature that considers the use of counterfactual analysis to elucidate the past, present, and particularly the future possibility of nuclear war (Lin 2021). It builds on Steven Webber's notion of "future counterfactuals" to construct imaginative yet realistic scenarios to consider the future possibility of a nuclear exchange. It highlights the critical role counterfactual scenarios can play in challenging conventional wisdom about nuclear weapons, risk analysis, war-fighting, and linear thinking.

Why counterfactuals are helpful analytical tools to view world politics? How can we optimize counterfactual reasoning to consider nuclear risk in the digital age? This framing of the problem highlights the critical role counterfactual scenarios can play in shaping perceptions and informing policy choices. In this way, the article contributes to our understanding of how thinking about the future – as much as policymaker's imagined past – is a critical function of nuclear war planning. In emphasizing the role of uncertainty, motivated and unmotivated cognitive bias (i.e. affect-driven vs. purely cognitive), and fundamental uncertainty in world politics, the article also contributes to the growing canon of literature that considers the growing risk of inadvertent and accidental nuclear war (Sagan 1993; Johnson 2021a).

The article is organized into four sections. The first describes why counterfactual scenario planning should be a core function of crisis decision-making and war planning. In addition to lessons of the past and other analogies (often with questionable accurately or relevance), imaginative critical thinking about the future with counterfactuals – which goes beyond net-assessment, mirror-imaging, and worse case-scenarios extrapolation of present trends – are powerful tools to challenge assumptions, expose bias, and weaknesses in the ways we frame research (empirical and theoretical) questions about the possibility of inadvertent and accidental nuclear war. How can counterfactual scenarios liberate policymakers from overconfidence, challenge assumptions, and expose bias?

Section two applies Weber's notions of a "counterfactual history of the future" to consider how we can best think about imagined futures – in contrast to conventional backward-looking counterfactual reasoning. It illuminates the choices policymakers face about the trade-offs associated with the intersection of DET's with nuclear weapons; during times of intense strategic competition, uncertainty, complexity, and mis-disinformation associated with today's emerging "information ecosystem" (social media, mobile communications, personal information feeds, and massive amounts of data from which people's interest and desires can be curated, etc.), which is characterized by the speed, volume, and scale of communications in orders of magnitude greater than in the past (Trinkunas, Lin, and Loehrke 2020). What criteria and processes can we use to counterfactual scenario planning to support this goal?

Section three unpacks Tetlock and Belkin's criteria to construct the "ideal-type" (Tetlock and Belkin 1996, 5) of scenarios for the purpose of crafting policy to account for future contingencies involving DET's and nuclear weapons. What might drive a particular situation or decision? What other factors might also affect this process? How far back into the "causes behind the causes" should we go in a scenario? This section highlights the differences between the future and past use of counterfactuals to shift our perceptions of possible futures and establishes a robust (i.e. falsifiable, realistic, and plausible) analytical framework to consider future contingencies involving emerging technology and nuclear weapons.

The final section applies the analytical framework to view three future counterfactual scenarios – an AI-enhanced cyber-attack on dual-use command and control networks; a third-party cyber "false-flag" operation; and disinformation operations to undermine confidence in a states' nuclear forces – to elucidate how emerging technology might spark escalation to nuclear conflict during an ongoing crisis between nuclear-armed adversaries (Johnson and Krabill 2020). The scenarios combine a driving force (the "plot") and process tracing analysis (See Figures 2, 3 and 4) The purpose of these alternative scenarios is to expose what is wrong with existing scenarios used by war planners and strategists and to advance an alternative approach (Favaro 2021).

Thinking forward with counterfactuals

Counterfactual scenario planning is a core function of crisis decision-making and war planning. Defense planners think counterfactually to construct scenarios about the potential impact of their policy choices and then retrospectively build "plausible worlds" to consider what could have happened if they or adversaries adopted alternatives. This analysis, however, rarely goes beyond extrapolating from present trends and mirror-imaging (Bronfenbrenner 2010). For example, the Cuban Missile Crisis became a subject of keen counterfactual investigation amongst scholars and policymakers, debating the importance of critical decisions made during this critical infection point in Cold War history (Allison 1971; Francis 2020; Lebow and Stein 1996). What other actors (adversaries and allies) think about the future will fundamentally determine the shape of their policy choices and thus crucially affect nuclear war planning, arms control, disarmament, and strategic stability dialogue proposals. Rigorous and empirically, and theoretical robust future counterfactual scenarios described in this article can illuminate fallacies such as "control" and inevitability (Considine 2021), the role of luck, overconfidence, as counterpoints to the conventional wisdom that denies (or underestimates) the past, present, and the possibility of nuclear war (Connelly et al. 2012; Knopf 2012; Pelopidas 2017). Decision-makers also often suffer from overconfidence in ways that risk escalating a crisis, which the scenarios in the final section will illustrate.

More broadly, political scientists (and by extension social science scholars) often fail to accurately anticipate the non-linear dynamics that can harbinger Schumpeterian-type (the theory of "creative destruction") technological change. Until AI machine learning algorithms can reliably supplant human predictive frailties in foreseeing the uncertainty associated with DET's (or AI "superintelligence") (Johnson 2004). counterfactuals are helpful tools to elucidate these processes. Whether and when we can move beyond the abstract (or even agree on) in developing self-learning AI systems – especially

unsupervised reinforcement deep-learning – that have predictive models capable of understanding their external physical environment and developing the requisite "theory of the mind" to interact with humans, creative counterfactuals encourage responsiveness to uncertainty and technological surprise.

Emancipation from the "Official Future"

When we apply counterfactual scenarios to illuminate situations, consider non-linear change, and challenge conventional wisdom, what are we doing? How might counterfactual scenarios liberate policymakers from overconfidence in their control of events, inevitability, challenge assumptions, and expose bias? And, if they can support this goal, how can they be optimized? Counterfactuals can help scholars interested in the intersection of technological change, nuclear risk, and world politics more generally to consider possibilities – seemingly remote or improbable alternative futures – and potentialities, thus reducing the degree of surprise and unpreparedness associated with change. This focus will also enable policymakers to recognize, respond to, and pre-empt technological inflection (or "bifurcation") points (e.g. next-generation algorithmic iteration or an adversary's operational concept) more effectively. When is the inflection occurring, by whom, and to what end, and are there trade-offs involved?

More fundamentally, counterfactuals can act as an effective defense against people's propensity for "hindsight bias" – elevating the probability of events once they have occurred and perceiving the past over-deterministically – which can lead to overconfidence and path-dependent technological deterministic mindsets (Fischoff 1975). Hindsight bias can also be viewed as a psychological coping mechanism for fundamental uncertainty, leading policymakers to overstate experts' confidence in the predictability of crises, accidents, and disruptive shocks (e.g. technological surprise, nuclear accidents, cyber supply-chain attacks, and foreign crises, etc.) (Hume 1999; Heidegger 1962; Rosenblatt et al. 1990). Moreover, this propensity is compounded by belief systems that allow people to remain unduly confident in the ability to predict a future marked by continuity and thus controllability – despite empirical evidence of discontinuities, shocks, and disruption in world politics, experts, were unable to either predict or control (Tetlock 2005). The proclivity of people to assume the future will closing map the past, can produce oversimplified, overgeneralized, and overdetermined insights from events (Jervis 1985b). In short, most people see the future as open and unpredictable but the past as overdetermined. If, for example, we could have asked people in 1913 the likelihood of averting what became known as World War I, we would receive very different responses that asking them this question in 2022.

A good case in point is the attachment of policymakers and scholars alike to the notion that "rational-based deterrence" (Lebow and Stein 1989) secured by the existential threat of nuclear war – grounded in the logic of mutually assured destruction (MAD) – will continue to ensure policymakers' (psychological and political) need for certainty and continuity in the ability to control nuclear weapons and thus prevent deterrence failure (George and Smoke 1974; Freedman 1989). To be sure, the fact that policy makers think their deterrence policies will succeed most of the time can lead to complacency that reinforcing "illusion of control" – allowing adversaries to manipulate and take advantage of this delusion (Jervis 1985a, 7). Empirical evidence from the Cold-War era, however,

paints a picture of repeated deterrence failure (i.e. deterrence failed to prevent the risks and situations it was designed to), nuclear "close-calls" (e.g. the Cuban Missile Crisis), and tautological justification for a policy that was essentially the product of counterfactual reasoning (Lebow and Stein 1996).

Counterfactuals are, therefore, valuable learning tools to challenge prevailing psychological bias (more on this below), enabling scholars to anticipate events and risks that in hindsight appear obvious, overdetermined, or at least, highly probable. Why, for example, did we fail to predict the collapse of the Soviet Union and the end of the Cold War? Despite several possible antecedents (e.g. declining productivity and birth rates, etc.), connecting principles (with well-established facts), and consistency of these principles with an established theory (see explanation below) to form a relatively high confidence forecast, the exponents of future counterfactuals at the time failed to dislodge the prevailing "official future" – i.e. a set of assumptions, bias, and perceived wisdom around which most bureaucratic and decision-making fora takes place (Kissinger 1993). In this way, counterfactuals differ from the more traditional retrospective-centric theory-testing used by international relations (IR) scholars – for example, policy-centric theories like the balance of power advanced by structural-realist IR scholars Stephen Walt and Kenneth Waltz et al (Waltz 1979).

In a world of growing complexity, automation, and information overload, filtered through human cognitive bias – and then funneled through and distorted by social media-fuelled filter bubbles and echo chambers – policymakers can use counterfactuals to play devil's advocate to objectively scrutinize both popular and unpopular hypothesis, which is a "an essential ingredient of a good detective, whether the end is the solution of a crime or an intelligence estimate" (Wohlstetter 1962, 16). Policymakers and political scientists are often taken by surprise and under/overestimate the impact in the event of technological change, in part because of a failure of divergent thinking – a tendency to not accord insufficient importance to the variety of possible futures that might yet occur, as well as possible pasts that could have been. Many predicted, for example, that chemical weapons would instantly and dramatically change the nature of warfare and deterrence after the British used poison gas during World War I. However, chemical weapons proved far less practical, impactful, disruptive, and relatively easy to defend against than conventional explosives (Brodie and Fawn 2010). More recently, US "network-centric warfare" did not prove to be the game-changing strategic innovation predicted by many military thinkers at the turn of the millennium (Rosen 2010).

Counterfactual scenarios can become tools of persuasion for decision-makers to contemplate one path over another, hedge against a latent (or improbable) risk, sidestep technological determinism, and heuristic thinking (anchoring, availability, and representativeness) (Tversky and Kahneman 1974), thus reducing false positives and negatives – i.e. perceiving risk where none exists and missing signs of possible risk for lack of looking. Decision-making that takes place in a complex, non-linear, and possibilist world – where the tape of history only runs once – which should persuade decision-makers of the value of counterfactual scenario planning to consider multiple plausible alternative scenarios of how past events and decisions might have otherwise occurred, and how future events might play-out (Schoemaker 1991). Furthermore, in domains where there is a paucity of empirical materials to satisfy basic epistemological criteria (i.e.

nuclear war), the prudence of this kind to manage risk and hedge against low probability events, where one failure is uniquely intolerable (i.e. accidental nuclear detonation), becomes a critical task (Weber 1996, 268)[4].

The creative thinking from counterfactuals can also reduce cognitive bias (discussed below) and misaligned perceptions associated, for instance, with notions of risk and control. Henry Kissinger described a proclivity of academics to believe in the notion of a "controllable world", causing deterministic reasoning that circumscribed policymakers' agency to influence events (Kissinger 1993). Kissinger's worldview gave prominence to the influence of individuals who could – through persuasion, inducements, and cajoling – be persuaded to change their minds. In turn, changes to perceptions can focus the minds of decision-makers, thus enabling them to create greater distance from often self-imposed (primarily structural, epistemological, and social-psychological) constraints, better anticipate non-linear change, and in turn, judge and respond to events with greater freedom and flexibility. As cognitive scientist Daniel Kahneman observed, counterfactuals can be powerful tools of persuasion and surprise by exposing hitherto unacknowledged tensions between explicit conscious beliefs and values and implicit, unconscious biases (e.g. hindsight bias and heuristics), double standards, contradictions, and assumptions (see below) (Kahneman 1995).

Consequently, the efficacy of counterfactuals is a measure more a function of their ability to alter perceptions (through theory, ideas, and debate) and ultimately behavior, rather than whether they can predict the future (Calvin 2006)[5]. As scholar Philip Tetlock notes, it is easy to appear correct for the wrong reasons, particularly when you have multiple alternative histories to hand (Tetlock 1992). A better post-mortem would include questions like What did the scenario over/underestimate, missed out, or miscalculate, and why? Were the results surprising, and if so, for whom, and why did they expect a contrarian result?

Future counterfactuals: "Counterfactual history of the future"

How different are counterfactual reasoning about the past compared to the future? Political scientist Steven Weber coined the term "counterfactual history of the future" to capture the notion that well-conceived future counterfactual scenarios share many similarities with traditional "counterfactuals histories of the past" (Weber 1996, 275–279). In this view, whether a counterfactual scenario is logically a question of where a question about where one temporally places an event (i.e. the counterfactual) at = T in relation to the past (i.e. what occurred) and the present. The underlying premise is that once one inserts a counterfactual into a historical sequence of events, the scenarios begin from that point along the timeline (see Figure 1). Put differently, at = T what has occurred ends, and the imaginary plot becomes the "logical future" (i.e. what occurs once the counterfactual is inserted). Viewed this way, all (past and future) counterfactuals can be considered *histories of the future*, even if they are inserted in what we would think of as past "calendar time" (Weber 1996, 277).

[4]This excludes, however, the use of counterfactuals to validate a control case of a null hypothesis.
[5]William Calvin argues that human's inability to foresee non-linear change reliably is due to the way we are neurobiologically hardwired in the nervous system.

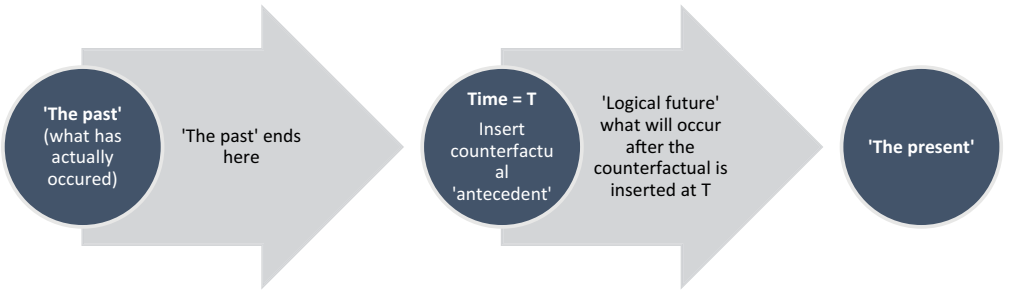

Figure 1. "Counterfactual history of the future".

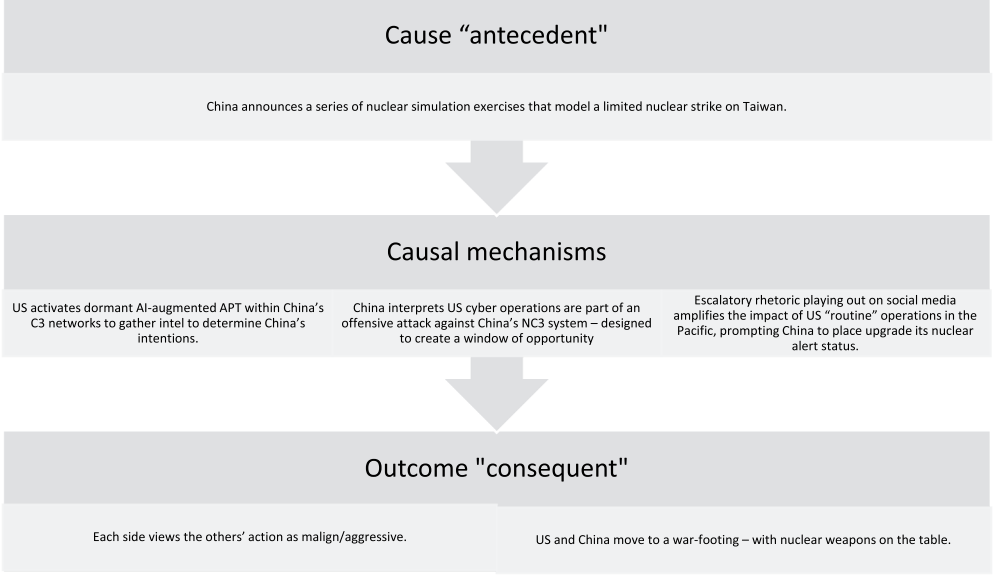

Figure 2. Process-tracing.

Regardless of where we situate the counterfactual that did not occur along the time continuum to the present day, it is about a possible fictional "future" that we can logically extend to and beyond the present. While "future counterfactuals" are an oxymoron, Weber's concept nevertheless proposes an interesting exercise that consists of exploring alternative outcomes to future scenarios where those designing or elaborating them *think they can predict the likely outcome*, or range of possible outcomes. In sum: whereas past counterfactuals consider imaginary changes of the past (antecedents) linked through a chain of logic to a different present (consequent), future scenarios confront a fictional world with a greater range and scope of imaginary alternative outcomes. The empirical implications of this distinction are discussed below.

A natural objection raised about this approach – in contrast to backward-looking counterfactual reasoning – is that while the past contains a limited data set to work off (i.e. the past is constrained by what we know happened), future permutations and combinations are, by definition, infinite. This objection can be addressed in three ways:

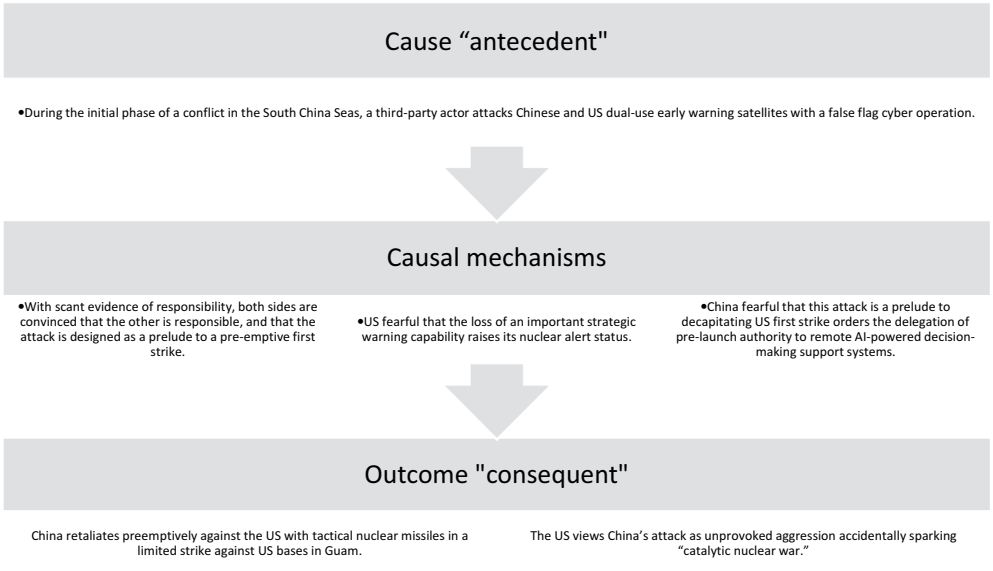

Figure 3. Process-tracing.

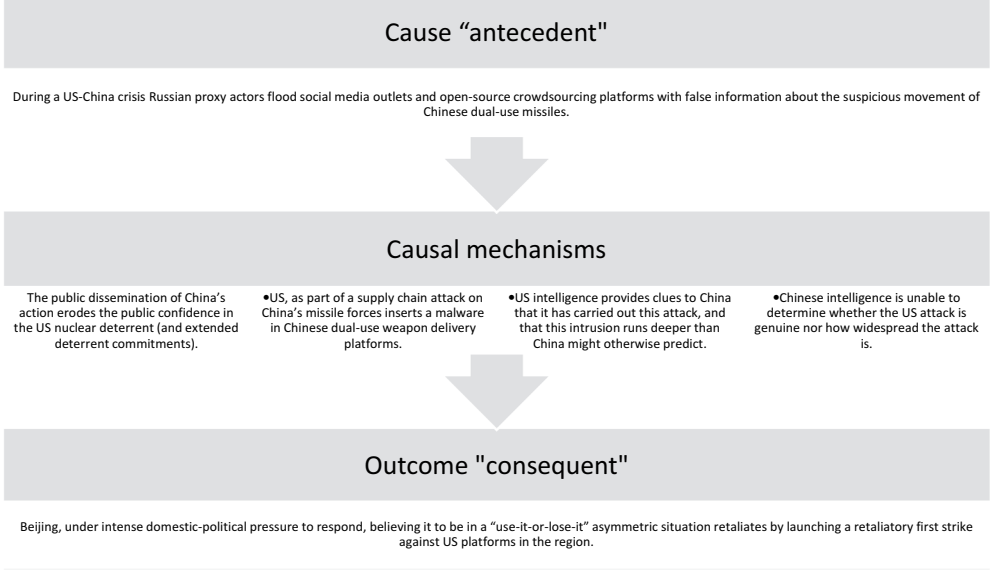

Figure 4. Process-tracing.

First, establishing valid and tested theories or set of criteria (i.e. “generalizability”) that increases the confidence in the internal consistency (or “contestability”) of an imaginary dataset of the future. Second, scenarios should be chosen that exist within the perimeters of current scientific reality. That is, new insights, models, and paradigm shifts fit in the Kuhnian “normal science” framework (or “Kuhn cycle”) governed by established

generalizations and laws[6] Third, when using counterfactuals to think forward, we must ensure that most of the observable implications (or "antecedents") that occur in our scenario planning are predetermined (i.e. already exist or our expected to shortly). DET's such as AI technology, robotics, quantum technology, cyberspace, for instance, fulfil this criterion (Kuhn 1962).

In framing the emerging technological problem in this way, we can avoid of technological determinism and consider alternative outcomes that illuminate the choices policymakers face today about the trade-offs associated with DET's, enabling them to hedge against unforeseen risk, improbable and unanticipated opportunities, and surprise. This analytical latitude and creative thinking are critical features of effective scenario planning, particularly during conditions of uncertainty, information overload, mis-disinformation, and great power strategic competition associated with today's emerging information ecosystem.

Should we necessarily have greater confidence in the more constrained retrodiction of counterfactual reasoning of the past? While it is axiomatic that social scientists' confidence in their retrodictions of past events is higher than future predictions, future uncertainty (i.e. infinite possible antecedents and consequences) can inculcate humility and reduce hubristic tendencies in our confidence of what we think we understand about the past. Because of the extraordinary complexity and ambiguity in devising plausible and falsifiable historical counterfactuals in international relations, the potential for randomly distributed minor causes (akin to "chaos theory") to be unduly amplified to explain more significant outcomes increases (Tetlock 1998, 32).

This psychologically bounded condition can mean that counterfactuals to view the past breakdown in the face of low probability and low-frequency events such as technological surprise and nuclear war (see below)[7] To avoid this fate, the article aims to strike a balance between parsimony (a single cause, future, or monotonic counterfactuals), and at the other extreme, the "all other things being equal" (a carte blanche to advance one's preferred causes and conditions). A failure to strike a proper balance between these two extremes risk hampering our ability to consider the role of luck, contingency, cognitive bias, and fundamental uncertainty in world politics. Reflection on these issues by policymakers who craft strategy, military doctrine, and operational concepts in the burgeoning information ecosystem is now a critical task.

Constructing future counterfactuals: Distinguishing scholarship from snake oil

Counterfactuals that foster creative thinking and encourage open minds – that challenge the "official future" – and force policymakers to consider uncomfortable discontinuities (or "plausible" or "realistic" futures) that may run contrary to the political and social Zeitgeist or prevailing theoretical canon (Lebow 2010, 44–45)[8]. What criterion and

[6]Kuhnian "normal science" refers to scientific progress advanced the most by occasional revolutionary explosions of new knowledge, each revolution triggered by the introduction of new ways of thought so large they must be called new paradigms.

[7]For example, it took several decades to discover the critical role misperceptions played in the 1962 Cuban missile crisis.

[8]"Plausible" counterfactuals should have a significant probability of resulting in the alternative outcome stipulated. More subjectively, "realistic" refers to elements of counterfactuals that do not contravene our current understanding of what is technologically, culturally, temporarily, or otherwise possible. This contrasts with so-called "miracle worlds" that are intrinsicality implausible.

processes should we apply to counterfactual scenario planning to support this goal? What follows is a description of the various criteria, methods, and processes that we can apply to counterfactual scenario planning to think about the future contingencies involving DET's and nuclear weapons[9]. The goal here is to yield "lawlike" generalizations – rather than scientific proof or Bayesian statistical confidence intervals – that foster creative thinking and open the minds of policymakers to consider apparent discontinuities from existing theories and assumptions (Vacher et al. 2018).

At this point, a caveat is in order. Oscillation exists amongst scholars about which criteria are most relevant (or irrelevant), conflicting or contradictory standards, and those that denounce the counterfactual project as epistemically and ontologically impossible. These criteria are all open to interpretation, and to date, no cookie-cutter universally accepted method of the counterfactual argument exists to apply to all situations and issues in world politics (Lebow 2015). These standards and processes represent an effort to establish compelling (or "ideal-type") counterfactual scenarios that command cross-disciplinary support and contribute to the broad social Science academic goals of logical consistency, falsifiability, parsimony, and have testable explanations. Thus, avoiding the fate of counterfactual critics, taking their cue from historian E.H. Carr, who argues that counterfactuals are arbitrary, speculative, self-serving, and non-falsifiable (Carr 1964). And scholars from rival cross-disciplinary schools talking past each other and at cross purposes. An attempt to map these attributes to construct robust future counterfactuals for scenario planning follows.

Future counterfactual building blocks

This section adapts Tetlock and Belkin's "best practice" criteria to distinguish between the plausible from implausible and insightful from wildly speculative counterfactuals to construct the "ideal-type" (Tetlock 1998, 5). of scenarios for the purpose of future scenario planning. Like most theoretical and ideological approaches, counterfactuals invariably lie at various points along the plausibility continuum (or Bayesian "subjective probability"). Because of this non-bifurcation, advancing "ideal" standards and processes for determining the utility of considering possible futures is a critical methodological exercise.

Specified antecedents and consequences and logical consistency between them

Like scientific experiments, counterfactuals should be designed to isolate pathways to a predetermined outcome (or dependent variable), should manipulate one cause (or independent variable) at a time. According to scholar Robert Jervis, in social systems (or "system theory argument") like world politics, it is impossible to hold "all other things being equal" when engaging in counterfactual scenarios, without generating ripple effects that value-basis of other potential causes in the interconnected historical matrix of interactions and complex social networks (Jervis 1976). It is extremely difficult, for example, to test (i.e. with a comparative method) the deterrence theoretical notion of

[9]These criteria and processes are adapted from a volume that surveys approach – both normative, epistemological, and cognitive – to view the role of counterfactuals in world politics.

situations resembling a "game of chicken" (Schelling 1960) – where an actor increases their chances of prevailing by signalling resolve to the other side – in which all the variables bar one are either the same or randomized is empirically problematic given the interconnections between them.

Unlike scientific experiments, however, counterfactual scenarios should not be too tightly wedded monotonically to independent and dependent variables; for example, if event X led to a nuclear detonation, then if X did not exist, the detonation would not have occurred. As a corollary, counterfactuals should not assume that one specific event or condition in the past (or future) can be altered and keep everything else constant (or "surgical counterfactuals") (Lebow and Stein 1996, 146). In other words, a moderate degree of specificity should be used in the construction of plausible scenarios, rather than imposing an unworldly rigidity and unambiguity on the relationship between antecedents and consequences.

As counterfactual skeptics rightly assert, manipulating antecedents that are so deeply embedded in a recursive system of a complex and interconnected "causal web" is misleading (Beebee 2019, 228–248)[10]. There is, to be sure, a long philosophical, epistemological, and metaphysical debate – going back to David Hume and Immanuel Kant's "First Critique" – about the nature of causation as an interpretative concept and method (Bennett, Saunders, and Stern 2019). The human brain cannot comprehend "all" other things, even if we know what they are. Moreover, in thinking forward with counterfactuals, this limitation is even sharper focus than backward thinking scenarios (Weber 1996, 285). The premise that "if cause X took on a different value, then Y occurs", for instance, does not pay due attention to the causal interconnectedness in complex social-political systems (Fearon 1996, 36–69). Critics of structural-realist scholar John Mearsheimer, for instance, argue that Mearsheimer's view that "*ceteris paribus*, war is more likely in a multipolar system than a bipolar one" glosses over the problem of what else would have to be different in a counterfactual post-World War II system in which multipolarity prevailed. For example, the effect of third power such as the UK, France, or China on the polarity dynamics (Mearsheimer 1990).

The logical consistency between the antecedent and consequence will depend on the scholarly consensus or shared faith – influenced in part by the prevailing IR theoretical canon – in the assessment of a particular connecting principle. Therefore, in designing a robust counterfactual, an essential first step is to identify the "driving forces" (a term that in contrast to independent variable implies a force pushing in a particular direction rather than what is known on one side of an equation) surrounding a particular event, problem, or decision – the building blocks of the scenarios' plot and its conclusion. Next, we need to consider and specify what else might change by manipulating a cause, the interaction of these altered variables, and how these changes may influence the possibility of the hypothesized outcome occurring. Other questions to consider include: What might drive a particular situation or decision? What other factors might also affect this process? And, how far back into the "causes behind the causes" should we go in a scenario? For

[10]The notion of "cause" is an intractable problem in all the sciences (social, biological, and physical). Prominent scholars, including philosopher David Hume, argued that this intractability results from the cause being a result of human mental activity and thus independent of the material world.

example, we can identify the increase in computing power, commercial interest, and expanded datasets in the genesis of artificial intelligence, but do we need to ask about the forces driving commercial interest?

Theoretical, empirical, and statistical consistency

Consistency of counterfactuals with well-established historical facts (or the "minimal rewrite rule") depends mainly on what scholars define as well-established and factual accounts of history. For example, in the absence of the assassination of Austrian Archduke Ferdinand First World War would not have occurred, which is a good case in point of a "minimal rewrite" counterfactual. In other words, the minimal rewrite rule is optimized when: a *high ex-ante probability* of a counterfactual antecedent is coupled with a *lower ex-ante probability* of real-world consequent (Levy 2015, 390).

Because of the competing schools of thought in IR theory that claim to account for the outbreak of war and the sustainability of peace (e.g. the balance of power, power transition, democratic peace, deterrence/spiral model, and neorealism), thus judgments about theoretical consistency are even more problematic. Critics of the predictive power of IR theories argue that in most cases, scholars' assertions rest on *ex post facto* (i.e. retroactively) and empirically questionable foundations to reconcile the validity of their theories with failing to anticipate real-world events. In the case of deterrence theory, for instance, predictive failures include the Cuban Missile Crisis, the end of the Cold War and the survivability of NATO, and the 1973 Middle East War. In this way, IR theories, much like other social science disciplines, are best thought of as an established means to enable scholars to apply past lessons as guides to the future – however fallible and contested they may be – rather than used as bulwarks against failed predictions and surprise (Tetlock 1998). Put differently, a theory stands or fails on its ability to generate propositions that are testable, valid, and falsifiable, rather than whether their assumptions accurately predict or explain empirical cases.

Cognizant of these limitations; therefore, an "ideal" scenario would manipulate one or more of the well-established facts – using a "well-established" IR theoretical lens – in a way that avoids ex-post-facto reasoning or manipulating other facts in the process (Weber 1996, 285). For example, scholars have used extended deterrence IR theory to argue that even if President Truman had not threatened Stalin militarily, the Soviets would have likely withdrawn from Iran. Robust counterfactuals, therefore, must clearly define and justify their theoretical (contested or otherwise) and empirical assumptions and describe what kind of evidence would likely increase the confidence in the validity of a hypothesis – or cause it to be rejected. This falsifiability requirement will help ensure realistic and plausible scenarios while avoiding snares in retroactive reasoning and non-falsifiable hypotheses.

The next step is to identify the elements of the scenario that are reasonably predetermined or "relatively certain" (Schwartz 1991, 117) from those uncertain and potentially significant – or "low probability events".[11] This step plays two crucial roles. First, it

[11]Peter Schwartz advances four examples of factors in world politics that are "relatively certain": slow-changing phenomena (demographics, climate change); restricted situations (budgets, electoral cycles); outcomes that are "in the pipeline" (emerging technology prior to deployment); and inevitable collisions (declining GDP and recession, trade wars and inter-state competition).

can help ameliorate the human cognitive propensity to overemphasize abrupt or discontinuous departures from normality – it is easier to "mentally undo" accidents or other events that constitute deviations from norm – and treat routine events as immutable facts (Hawkins and Hastie 1990). There is, however, much empirical evidence of sudden and unexpected departures from the status-quo (regime change, revolutions, assassinations, and war, etc.), attracting much retroactive counterfactual thinking (Ferguson 2014). Second, specifying what is relatively sure from what is not can help policymakers consider what *other things might not be equal* in a scenario and how we might know. Thus, taking one step beyond "all other things being equal" approach.

In many situations in world politics, where empirical data is scarce (nuclear war), we have few validated empirical laws to draw on (international relations), and competing theoretical schools prohibited consensus (offensive vs. defensive realists or structural realists vs. constructivists), statistics can help fill in the gap in what could have happened (or might occur). Such as if an event or decision had occurred or been omitted – that is, the importance of negative evidence or "dogs that do not bark".[12] Game theorists, for example, use statistical reasoning – such as comparative statistics and probabilistic models – to test and design counterfactuals (Collins, Hall, and Paul 2004, 50). In game-theoretic models of interactive decision making, statistical models explicitly compute what would happen if actors made different choices, how other actors would strategically respond, and the payoffs that actors would gain (or loss) for every possible combination of choices[13]. In robust counterfactuals – like empirical and theoretical laws – connecting principles of antecedents and consequences should align with relevant "well-established" statistical generalizations, for example, base rates and patterns of covariation. According to Robyn Dawes, for example, counterfactual inferences are only plausible if they are inculcated in an established statistical system supported by "reasonable" empirical evidence (Dawes 1988). Similarly, Hume argued that inference cannot be a matter of "demonstration" or *a priori* inference – we cannot use reason to infer from an event alone what might happen next (Beebee 2019).

In contrast to other domains such as the stock market, games such as chess and Go, and sports, however, well-established generalizations in world politics (war, peace, crisis management, the impact of technological change, regime type, and strategic bargaining, etc.) are notoriously difficult to model. Therefore, the use of statistical studies to construct counterfactual scenarios of technological surprise, warfare, and international relations more broadly defined, is problematic for ontological, psychological, and normative reasons. Statistical models and logical reasoning are unable to capture the qualitative discontinuities and uncertainties that shapes the complex interactions between states (the rules of the game, international institutions, and perceptions etc.).

In most cases, how policymakers behave will depend on a multitude of non-linear causal conditions (e.g. lessons from past cases, theory-driven, regime type and domestic-political constraints, alliance structures, perceptions of the military balance, and technological change), which will vary considerably temporally (the political, cultural, organizational context) and spatially (lessons from past cases, shifting perceptions and mindsets)

[12]The problem of explaining causation by and of omission (i.e. explaining what is needed to prevent an event from occurring or to permit events to occur) is an area of continued research.

[13]The decision that is not made, and the sequence of decisions that would have followed from them, are defined as "off the equilibrium path" – the criterion of subgame perfect equilibrium.

(Beckert 2016). Policy makers tend to be heavily influenced by recent events, which actors have experienced first-hand, and events that occurred when they developed political awareness (Jervis 1976). Counterfactual reasoning can help to reduce the risks of forcing new empirical data into pre-existing theoretical paradigms.

According to sociologist Jens Beckert, how we anticipate the future is very different from economists' – such as Cold-War era neoclassical economics and rational choice theorists at the RAND Corporation – project rationale preferences along with a set of knowable probabilities. Beckert argues that the future is a function of fundamental uncertainty instead of building on Frank Knight and Kenneth Boulding's emphasis on subjective judgment and "omnipresent uncertainty" in contemporary risk studies (Beckert 2016). It prompts actors, as a corollary, to resort to beliefs, fictional narratives, and an imagined conceptualization of the future that allows them to sustain the illusion of linear temporal continuity (Knight 1921).

While falsifiability may be a cornerstone scientific method for testing a hypothesis, believability (or plausibility) constitutes the hallmark of an effective narrative or plot (Popper 2005). Consequently, the role of uncertainty needs to be considered alongside the human psychological need for forms of continuation of the present – the fictional scenarios below explore this synthesis. Lee Clarke observed that "we need to think in terms of chances and odds and likelihoods. But we shouldn't concentrate so much on probabilities that we forget the possibilities" (Clarke 2010, 41). The universe of possible future cases and permeations of these interactions is infinite; thus, thinking forward with counterfactuals can help reduce the risk of false certainty.

An additional obstacle in statistical studies for counterfactual scenarios – which is borne out in the cognitive psychology and experimental literature – relates to human bias. Namely, people are notoriously flawed intuitive statisticians, whose biases are frequently laid bare in detecting and using covariation data – the measure of how two random variables in a data set will change together – which impairs our ability to determine the plausibility (or "reasonableness") of counterfactuals (Nisbett and Ross 1980). While this cognitive fallibility can be ameliorated by improving the accuracy of covariation estimates (e.g. including omitted data), this solution does not resolve the problem of peoples' propensity to accept false counterfactual claims (i.e. false positives caused by Type 1 errors) and dismissing true ones (i.e. false negatives caused by Type 2 errors) (King, Keohane, and Verba 1994)[14].

In other words, the problem here is not with counterfactuals *per se* but instead deeper cognitive-psychological issues that can affect the choice of variables. As Hume posited, the "ideas which we form" to create causal relationships, because of our cognitive limitations, are inherently "imperfect" (Beebee 2019, 243). Exposing these (oft-implicit) biases that people carry with them – informing and shaping decision-makers' "official future" – driving the tendency for the certainty of hindsight, underestimating the role of luck, and overconfidence is thus a key goal of counterfactual scenarios. Because of the complexities and uncertainties involved in world politics, statistical reasoning probability

[14]Other biases related to statistical reasoning include those caused by non-random selection, omitted and confounding variables, which can also upend robust statistical inference.

estimates provide a very crude measure of the infinite of possible future events that might arise inside-outside and on the margins of interacting antecedents which link to potential consequences.

Given the obstacles and limitations described, the counterfactual scenarios illustrated in this article will not include probability estimates. Instead, they will construct disciplined counterfactuals – grounded inconsistent and well-established empirical and theoretical generalizations – to illuminate plausible scenarios involving the intersection of DET's in the "nuclear enterprise", challenging the "official future" of policymakers[15]. That is, wedded to outmoded and obsolete Cold War-era theoretical (i.e. rational-deterrence) mindsets, shifting perceptions, so that we can reduce the element surprise from technological change.

Lawlike generalizations to anticipate the future

How can we determine the extent to which the lawlike generalizations we have constructed are robust enough to support projections into the future? Nelson Goodman's concept of "projectability" is beneficial towards this end (Goodman 1983). Goodman distinguishes between coincidental generalizations – that just happens to be a specific time and place – and robust lawlike generalizations (e.g. oxygen is a necessary but not sufficient for fire) that hold up to a range of tests and allow projection into the past and future[16]. Thus, whether a generalization is deterministic or probabilistic, or bounded or unbounded by moderator variables, the litmus test is essentially the same, namely, the ability of a generalization to predict or project what is likely to occur in the future in *a priori* unobserved situations. For instance, if a hypothesis is true, what else should also be true or observable?

As a corollary, and akin to philosopher Karl Popper's "falsification principle", a counterfactual scenario should comprise a lawlike generalization that yields a falsifiable prediction (Popper 2005). For example, the hypothesis that "all swans are white" can be falsified by observing a black swan. However, constructing compelling and robust counterfactuals to look to the past and future are not symmetrical. When we use counterfactuals to look backward, there is a strong proclivity towards using antecedents that make a situation appear part of a predetermined and inexorable causal relationship, which risks overstating the role of chance – that is, the subjective probability assessment of how likely it is that a particular outcome might occur. Political scientists generally begin (and often end) their research by locating historical situations and then retrospectively searching for factors that appear correlated to the outcome. This process tells us very little about either the relative frequency of a particular outcome or proves that the variables used are a necessary causal factor; or that they may not be found in situations

[15]The "nuclear enterprise" refers to the complete range of activities, capabilities (nuclear and non-nuclear), and operations that directly or indirectly interface with nuclear weapons, including: production, acquisition, operations, organization, and strategy.

[16]Most political science generalizations are neither coincidental nor robust lawlike; instead, they are most contingent and bounded by variables (i.e. x causes y; x also causes z) or statistical generalizations (i.e. x increases/decreases the probability of y).

where the result is different (Jervis 1989). This approach characterizes, *inter alia*, research on alliances, arms races, the procurement of weapons, civil unrest, causes of war, and deterrence failure.

Retrospective causal relationships are often poor predictors of the future[17] This asymmetry indicates that counterfactuals in world politics can fail not because their core premise is false, but rather because of the inherent complexity interactions between causal variables that – like "chaos theory" – generates an infinite potential for minor causes in low-probability and low-frequency events (i.e. nuclear wars) being amplified into larger effects. Many IR scholars, for example, consider World War I to indicate a broader linear causal chain of related events and decisions – or "Humean causation". The theory-building of many IR scholars' centers on the value of structural (or systematic/linear) factors (i.e. between units of the international system) while paying lip-service to apparent transitory importance of non-structural (or non-systematic/non-linear) factors such as the role of luck, chance (i.e. accidents, third-party intervention, and technical failure), and causal bifurcation. Consequently, many predictions underplay (or discount) the role of complexity, by implicitly assuming that each casual factor exerts an independent influence on events, rather than interacting with other factors to generate non-linear outcomes. This assumption – which is also demonstrated in theorizing on critical junctures, positive feedback, and path dependency – can in part be attributed to a cognitive bias that connects perceptibly significant causes to large events (McCloskey 1990).

Several cognitive psychologists argue that low-probability and low-frequency (or "bolt from the blue") events – accidents, mechanical failure, technological surprise etc. – are precisely the class of events that are most likely to attract counterfactual thinkers. The literature on "normal accidents" suggests that eventually these kinds of events always will happen (Perrow 1984). Well-designed counterfactuals that illuminate surprise "bolt from the blue" and other long-short close-call events can recruit a related cognitive disposition to good effect. Namely, the tendency to judge a scenario as plausible if an event is depicted in vivid and narrow terms, leading people to upgrade their confidence in the probability of an event (e.g. accidental nuclear war) contingency – their perceptual lens of what they consider possible or rational outcomes can be widened when exposed to scenarios with divergent and improbable outcomes.

Finally, the idea of "temporal proximity". That is, the confidence we have in the validity of a counterfactual scenario as a function of the temporal distance and length of the causal chain from antecedent to consequent. In other words, the "distance" between actual and plausible worlds. The greater the distance and the longer a scenario's causal chain, so the greater the likelihood for bifurcation in the causal chain, for other events and conditions to intervene and shift the future onto a different trajectory, in turn, making the outcome less predictable and uncertain. An insurmountable obstacle, therefore, to make anything other than short-term predictions. According to historian J.D. Gould, "almost all of the obstacles to accurate prediction grow, some of them exponentially, as the time horizon is extended" (Gould 1969, 199–200). Echoing this view,

[17]For example, in the aftermath of the 1970 Mexico City passenger airline crash, the causal retrospective story that FAA investigators constructed – listing many plausible antecedent causal variables such as weather conditions, smog, radio malfunction, and fatigue, etc. – did not help the FAA to predict future accidents.

historian Ian Kershaw argues that scholars should use "short-range counterfactuals" rather than engage in an "intellectual guessing-game of looking into some distant future". (Kershaw 2007, 6). The scenarios used in this article adhere to James Fearon's "temporal proximity criteria" that "only where the [past or future] counterfactuals involve causal mechanisms and regularities that are *well understood* and that are considered at a spatial and temporal range *small enough that multiple mechanisms do not interact*, yielding chaos" (emphasis added) (Fearon 1996, 39–67).

Future counterfactual scenarios

This section applies the methods and processes described to three future counterfactual scenarios that challenge prevailing assumptions (overconfidence in the controllability of nuclear crises), bias (hindsight bias and heuristics), and prepare policymakers to hedge against unforeseen risk, the improbable, and unanticipated technological surprise. The scenarios are constructed as illustrative rather than empirically exhaustive. They are designed as reflective tools to expose weaknesses in the prevailing "official future", revealing potential blind spots or shortcomings in how we frame our research questions to consider nuclear risk anew in the digital age. While the scenarios viewed in isolation may appear improbable or preventable, taken together, they demonstrate how the interaction of a range of technologies with nuclear weapons might exacerbate crisis stability and increase the risk of nuclear escalation.

Towards this end, the scenarios combine a driving force (the "plot") and process tracing analysis. Research puzzles include: Can a causal chain be traced back to underlying causes and enabling conditions? Does a demonstrable (or incontrovertible) catalyst exist? Are multiple causal chains involved? If so, to what degree are they independent (or dependent) of each other? The more independent, the more probable the outcome can be understood as the result of confluence – that is, multiple independent causes combine to produce an event that might not otherwise have occurred (Lebow 2015, 409). The counterfactual scenarios are premised on four assumptions: (1) crisis conditions involving two nuclear-armed adversaries; (2) an intense "security dilemma" operating within the dyad (Butterfield 1951; Herz 1951; Jervis 1978); (3) the existence of military (both capabilities and information) asymmetry between adversaries; and (4) the technical feasibility of the operations and capabilities illustrated – either currently deployed or being developed. These provisional assumptions, together with the causal claims described below – and without pre-existing empirical evidence – allow us to construct rigorous scenarios that meet the positivist standards established earlier in the article. These foundations, in turn, will require modification evaluation rather than testing as new data, competing scenarios, or assumptions emerge[18]. In other words, this process is a form of social science process tracing (See Figures 2, 3 , and 4) to support a hypothesis with observable implications thinking forwards, rather than to the past.

Hypothesis: *Nuclear rivals are more likely to use military force against each other in ways that could escalate (inadvertently or accidentally) a crisis to a nuclear conflict (Y) than they might have if they did not possess disruptive emerging technology (X)*

[18]Space does not allow for exploring a range of plausible alternative scenarios and "wild card" events and the uncertainties that might arise from the possible combinations of them.

Scenario (1): AI-enhanced cyber-attack of dual-use command and control systems during a nuclear crisis

Driving forces: The "Plot"

China announces a series of nuclear simulation exercises that model a limited nuclear strike on Taiwan. The stimulation includes moving the PLA's Rocket Force – responsible for China's nuclear weapons – mobile missiles out of their garrisons and increasing the resiliency of China's dual-use (i.e. supporting both its nuclear and conventional capabilities) command, control, and communications (C3) systems (Acton 2018; Johnson 2019).

The United States, to ascertain Chinese intentions in conducting these "non-routine" exercises during a period of heightened Cross-Strait's tension, activates dormant an AI-augmented advanced persistent threat (APT)[19] clandestine cyber-attack within China's C3 networks to gather intel – to determine whether the stimulations are an exercise or part of a clandestine preparation for an actual first strike (Buchanan and Cunningham 2020). China's C3 network resiliency efforts (part of its simulation exercise readiness) enable it to detect the US's APT intrusion – which hitherto remained undetected – and thus assumes US cyber operations are part of an offensive attack against China's nuclear command, control, and communications (NC3) system – designed to create a window of opportunity for a larger strategic strike (Blair 1993)[20]. Thus, precipitating a major crisis where only a minor one existed before[21]. Therefore, in the absence of these enabling novel technological advances and techniques (APT, cyber counter-measures, dual-use technology), this causal escalatory chain of events would be technically unfeasible and thus operationally implausible.

Because the discovery of the US cyber intrusion coincides with escalatory rhetoric between the opposing sides diplomatic spokespersons playing out on social media, and corresponding provocative US freedom of navigation and a series of aircraft within Taiwan's air defense identification zone (ADIZ), China shifts its nuclear status to high alert. Amid recent Chinese aggressive disinformation campaigns in Taiwan and a nuclear modernization program, Washington is unassuaged by Beijing's statements that these exercises are merely a simulation.

Its clandestine cyber operations are designed to determine China's genuine intentions from the US perspective Green and Long 2019/20). That is, the US assumes China's exercises are malign, and given its upper-hand vis-à-vis China, feels less of an imperative to back down or refrain from escalation. Therefore, both sides are incentivized to take a more aggressive military posture, but for different reasons. The US views China's "mere" exercise as mobilization to a war-footing stance, and China views US penetration of its NC3 network as a deliberate attempt to comprise these capabilities. In sum: in the absence of these enabling novel technological advances and techniques (APT, cyber

[19]The cost of tools used to create malicious documents depends heavily on whether the malware can persist on the target system undetected by antivirus software.

[20]Technological advances in AI technology and cyber capabilities, coupled with the increasingly commingled nature of the state's nuclear and conventional command and control systems, have enabled solutions to overcome the robustness of permissive action links and increase these vulnerabilities systems.

[21]China would be unable to know what the US intentions were *before* the operation was detected (i.e. espionage or an offensive cyber-attack), and thus it would likely assume the worst.

countermeasures, dual-use technology), all things being equal, this causal escalatory chain of events would be technically unfeasible and thus operationally implausible, thus supporting the hypothesis.

Scenario (2): Third-party cyber false flag operation

Driving forces: The "Plot"

During the initial phase of a conflict in the South China Seas, a third-party actor (terrorist, state proxy, or other criminals, etc.) launches a "false flag cyber operation" (e.g. data manipulation, social media flooding with bots, a spoofing attack, or other forms of deception), to realize their "apocalyptic world views", (Fitzpatrick 2009; Ferguson and Potter 2004; Forest 2012) against Chinese and US dual-use early warning satellites – providing both sides with warning of a nuclear attack on their homelands – which is untraceable and thus appears to both states as originating from the other (Lin 2012)[22]. No evidence exists to disprove these claims, and both sides deny responsibility (Shao et al. 2018)[23]. Each side is convinced that the other is responsible, and the attack is designed as a prelude to a pre-emptive first strike[24].

The US, fearful that the loss of a critical strategic warning capability puts its launch-on-warning nuclear forces at risk, raises its nuclear alert status (Gamel 2017)[25]. As a de-escalatory tactical measure (to signal resolve), the US launches a cyber malware attack against China's conventional C3 network. The attack accidentally disables one of China's NC3 nodes, which Beijing views as intentional and malign. China, fearful that this attack is a prelude to the US attack to sever the control of its nuclear forces from the CCP leadership, orders the delegation of pre-launch authority to remote AI-powered decision-making support systems (Jia and Zhou 2016).

Lacking sufficient evidence of who is behind the attack and under intense domestic political pressure to respond, China launches a preemptive limited nuclear strike against US bases in Guam. The US views China's attack as unprovoked aggression, sparking an accidental "catalytic war" (Kobe 1962; Johnson 2021b)[26]. Once both sides consider conflict inevitable, the security dilemma logic of seizing the tactical first-mover advantage – before the other side can fully execute their war plans can create self-fulfilling spirals of escalation. In sum: like the first scenario, without the confluence of the complex

[22]In a conventional kinetic conflict, an actor may attempt the increase the effectiveness of its short-range missiles through interfering with or manipulating an adversary's early-warning satellites (especially high-frequency communication satellites), for example, a cyber-attack on its satellites ground links. However, an attack like this could lead the victim to assume that its long-range strategic forces were under attack.

[23]While a cyber false flag operation would not require particularly sophisticated technical expertise – at least compared to the level of scientific and military infrastructure required to develop nuclear weapons – it would need the organizational know-how to collect and parse intelligence and conduct clandestine activities.

[24]Given the pivotal role of command and control, a state's NC3 systems would likely by the targeted to degrade these capabilities during the initial stages of a conflict.

[25]In 2017, for example, US soldiers and their family members in Korea were subject to a text notification with fake evacuation orders for non-combatants issued for the Korean Peninsula. Though this attack was quickly debunked, a third-party attack like during an ongoing crisis could prompt a state to order a pre-emptive strike, assuming that the evacuation was a prelude to conflict initiation.

[26]The notion of "catalytic war" is the possibility that a third party's actions spark a nuclear war between the two nuclear-armed opponents.

and interdependent contingencies enabled by emerging technologies (malware, AI decision support tools, NC3 automation) described here the recourse to military force would, all things being equal, likely be significantly less – and supporting the hypothesis.

Scenario (3): Information warfare operations to undermine confidence in a states' nuclear forces

Driving forces: The "Plot"

During a US-China crisis, Russian proxy actors flood social media outlets and open-source crowdsourcing platforms with false information (satellite imagery, 3D models, Twitter feeds, or geospatial data, etc.) about the suspicious movement of Chinese dual-use (nuclear and conventional capable) DF-26 intermediate range nuclear road-mobile launchers, able to reach US military assets in the Pacific[27]. Once the Russian operations (e.g. deepfakes and bots) went viral, neither US human nor machine operators were able to limit the fall-out from the attack eroding the public and policy-makers confidence – a psychological as well as a technical phenomenon – in the US's extended deterrence commitments in the Pacific (Zuboff 2019)[28].

The US, as part of a "supply chain attack" (Schneider and Sherman 2021)[29] on China's missile forces – unable to determine with confidence the veracity of this information and with mounting public pressures to respond – inserts malware in several Chinese dual-use weapon delivery platforms. Against the backdrop of an escalating crisis – and based on misinformation "false positive"[30] – US intelligence provides serendipitous clues to China that it has carried out this (unprovoked) attack, allowing it to discover these system vulnerabilities[31]. The United States also informs Beijing that these insertions have infected other Chinese missile platforms – including China's dual-use DF-26.

The Chinese leadership is unable to determine whether the US attack is genuine – and not part of a disinformation campaign to signal resolve to China and reassure its regional allies – how widespread the US penetration is, or whether the United States gained access to additional vulnerabilities that it did not reveal (e.g. NC3 networks), and thus how to respond. Under intense domestic political pressure to respond, Beijing believes it is in a "use-it-or-lose-it" situation, retaliates by launching a retaliatory first strike against US

[27]From an intelligence standpoint, nuclear solid-fuel missiles and tracked TELs reduce ISR systems' ability to detect signs of launch preparation. Solid fuel also increases the speed of launching missiles and reduces the number of support vehicles to support an operation.

[28]Facebook, for example, uses its algorithms to anticipate human behavior to create "prediction products" that make people easier to manipulate – i.e. profile and micro-target their users to sell more advertising space. This capability was allegedly used to manipulate public perceptions during the 2016 US presidential election and the United Kingdom's referendum on membership of the European Union.

[29]Supply-chain attacks on complex and interdependent systems (e.g. the "colossal" Kaseya ransomware supply-chain attack in 2021) can render parts or the entire chain vulnerable before it is put into use.

[30]Alternative outcomes from this fictional scenario are, of course, possible. For example, counter-AI systems might uncover the leak's source or false nature before it can do severe damage. State A might also assure State B through backchannel or formal diplomatic communications of this falsehood. While social media platforms have had some success in slowing down users' ability to orchestrate manipulative and dangerous campaigns, once these operations (e.g. deepfakes and bots) go viral, the ability to curtail them becomes inexorably problematic – for human operators or machines.

[31]If an actor can demonstrate that it has successfully comprised one or more weapon systems, then a claim that it has also comprised multiple systems on a leader's decision-making might be akin to the reality of such an operation. That is, the destructive potential of such an attack could severely affect a leader's confidence and risk appetite.

platforms in the region[32]. In sum: the coalescence of a range of interconnected information technologies (satellite imagery, 3D models, geospatial data, social media mis/disinformation, and deepfakes), and the escalatory causal chain of events that followed, would, all things being equal, be unlikely to occur in the absence of these enablers thus supporting the hypothesis.

Conclusion

This article builds on the notion of "future counterfactuals" to construct imaginative yet realistic scenarios to consider the future possibility of a nuclear exchange. It highlights the critical role counterfactual scenarios can play in challenging conventional wisdom about nuclear weapons, risk analysis, war-fighting, and linear thinking. It argues that future counterfactuals are an important analytical tool to supplement traditional backward-looking counterfactuals, lessons of the past (or "nuclear learning") (Knopf 2012) and other analogies, to construct imaginative scenarios that challenge conventional wisdom ("illusion of control", rationality, the role of luck, inevitability), assumptions (rationality, path-dependency, over-determinism, control), and human bias (hindsight bias, heuristics, and attribution bias).

While the article makes a timely contribution to the scholarship about the use of counterfactuals in security, strategic, and nuclear studies, it is not merely an intellectual exercise. Drawing from multi-discipline best research practice criteria, the article advanced a methodology and causal mechanism – to construct robust (empirically and theoretically), rigorous (falsifiable, realistic, and plausible), and analytically parsimonious – yet going beyond "all other things being equal" – future scenarios to consider the risk of inadvertent nuclear conflict in the digital age. The research threads unpacked in the article also have significant implications for the nuclear policymaking community. Namely, the critical role imagined futures – which go beyond extrapolation and mirror-imaging – can play in shaping perceptions and informing policy choices. By giving prominence to the role of uncertainty, cognitive bias, and fundamental uncertainty in international relations, the article also contributes to the nascent literary canon about the risk of inadvertent and accidental nuclear war.

The counterfactual scenarios demonstrated that multiple causal chains would likely be involved in a future crisis dynamic between a nuclear-armed adversarial dyad – rather than evidence of a demonstrable or incontrovertible catalyst. Moreover, the multiple casual chains illustrated in the process tracing analysis (dual-use C2 technology, public confidence in nuclear deterrence, third-party information operations, preemption doctrine, AI-enhanced cyber capabilities, pre-launch authority, and social media-fueled escalatory rhetoric, etc.) were demonstrably interdependent and mutually reinforcing.

The process-tracing analysis also highlighted the critical role of human psychology, connecting the antecedents (i.e. counterfactuals) to the consequent (i.e. outcome) through the casual chain of events. Research on human cognition demonstrates peoples' tendency to deviate from rational decision-making pathways when faced with situations

[32]The United States and Russian nuclear doctrine maintain the option for counterforce operations to limit the damage it would suffer from a nuclear exchange or believe that the other side might launch a counterforce attack. Observers have also debated whether India and China are moving in the same direction.

combining complexity, uncertainty, stress, and risk (Kaplan, Wanshula, and Zanna 1993; De Dreu 2003). During times of crisis this propensity can influence the kinds of information leaders to give credence and prioritize, which often leads to a tendency to explain the behavior of others (i.e. adversaries) as the consequence of their inherent disposition and the same behavior in oneself as the result of factors out of their control – known as "fundamental attribution error" (Rodd 1977). Threats, risk, chance, and control are not entirely rational cognitive processes; they are as much (if not more so) psychological and emotional constructs, particularly during time of crisis. Recent neuroscience research, based on brain lesion studies, clearly demonstrate that emotion is necessary for any form of rational decision-making to take place.

This cognitive disposition, coupled with an intense security dilemma, can help us understand the causal pathways and leadership decisions made in the scenarios[33]. These findings also give credence to the literature on deterrence theory that critiques the core assumptions underpinning rational-based deterrence – namely, rationality, freedom from domestic-political constraints, the ability to identify aggressors from defenders, and challengers are risk-accepting maximizers. A noteworthy recurrent theme in the scenarios was the effect on crisis stability of dual-use systems used in conjunction with clandestine technological operations (whether defensive or offensive) had on both sides perceptions, attitude to risk, and confidence – both in attributing the intentions of the other and assuming their motivations are apparent and immutable. The process tracing suggests that as a corollary, the cognitive-psychological impact on decision-makers of these operations (especially deploying AI-enhanced cyber capabilities) was equal, if not more significant, than the actual effect of the operations per se.

The feasibility (or realizability) and impact (on crisis stability) analysis revealed additional research threads, including: 1) the indistinguishability of offensive and defensive DET-augmented capabilities and operations (especially in cyberspace); 2) the inadvertent escalation risk associated with commingled (or entangled) nuclear and non-nuclear capabilities (mainly C3 and ISR systems); 3) temporally, the importance of the geopolitical context in shaping perceptions, and thus, informing policy; 4) destabilizing effects of offensively-oriented military doctrine (launch-on-warning, preemption, etc.); and 5) the amplifying effects of the information ecosystem (social media, deepfakes, bots, and disinformation, etc.) on public opinion, and in turn, decision-makers perceptions (Garfinkel and Dafoe 2019).

Three lines of research inquiry flow from the article's findings. First, primary research could be conducted (interviews with experts, declassified wargaming exercises, table-top events, etc.) to ascertain how prevalent the "official future" is. How common are dismissals of scenario lessons? What kinds of cognitive bias participants exhibit, and do adversaries and allies share these attitudes? And which kinds of emerging technology are most (and least) likely to spark inadvertent escalation? Additionally, AI machine learning techniques – modeling, simulation, and analysis – may complement counterfactuals and low-tech table-top wargaming simulations to identify contingencies under

[33]These outcomes of the scenarios might unfold differently depending on alterations to various observable implications such as changes in leadership, regime type, the political Zeitgeist, technological breakthroughs, and the attitude and reaction of allies/partners, etc.

which "perfect storms" might form; not to predict them but rather challenge conventional wisdom, highlight bias, and inertia, to highlight, and ideally, mitigate these conditions (Davis and Bracken 2022).

Second, scholars should continue archival work on historical cases of nuclear near misses and employ counterfactuals (past and future) to explore contingency, the role of luck, and the Knightian relationship between risk and uncertainty. The goal here is to develop more sophisticated ways to consider future risk and avoid survivability bias – it is difficult to conceive of something abstractly such as a catastrophe or apocalyptic danger (Anders 1962).

Third, historians, social scientists, industry experts, and policymakers should collaborate earnestly, and with open minds, in the use of counterfactual scenarios to probe actors imagined past and future to expose and objectively understand our blind spots in the risk of nuclear war caused by a lack of imagination or a refusal to believe in its possibility – whether the result of myopia, control delusions, survivability bias, or the fear of potential dangers ahead. This line of effort, to be sure, could be explored to think critically about the possibility of a range of seemingly improbable existential threats such as climate change, pandemics, and artificial general intelligence (or "superintelligence"), to name a few.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Notes on Contributor

Dr. James Johnson is a Lecturer in Strategic Studies at the University of Aberdeen. He is also an Honorary Fellow at the University of Leicester, a Non-Resident Associate on the ERC-funded Towards a Third Nuclear Age Project, and a Mid-Career Cadre with the Center for Strategic Studies (CSIS) Project on Nuclear Issues. He is the author of *Artificial Intelligence and the Future of Warfare: USA, China & Strategic Stability*. His latest book project with Oxford University Press is entitled *AI & the Bomb: Nuclear Strategy and Risk in the Digital Age*.

ORCID

James Johnson http://orcid.org/0000-0002-5203-8583

References

Acton, J. 2015. "Hypersonic Boost-Glide Weapons." *Science & Global Security* 23 (3): 191–219. doi:10.1080/08929882.2015.1087242.

Acton, J. 2018. "Escalation through Entanglement: How the Vulnerability of Command-and-Control Systems Raises the Risks of an Inadvertent Nuclear War." *International Security* 43 (1): 56–99. doi:10.1162/isec_a_00320.

Allison, G. 1971. *The Essence of Decision. Explaining the Cuban Missile Crisis*. New York: Little Publishing.

Amis, M. 1987. *'Thinkability' in Einstein's Monsters*. New York: Harmony Books.

Anders, G. 1962. "Theses for the Atomic Age." *The Massachusetts Review* 3 (3): 497.

Beckert, J. 2016. *Imagined Futures. Fictional Expectations and Capitalist Dynamics*. Cambridge MA: Harvard University Press.
Beebee, H. 2019. "Hume and the Problem of Causation." In *The Oxford Handbook of Hume*, edited by Russell, Oxford University Press. 228–248.
Bennett, C., J. Saunders, and R. Stern. 2019. *Immanuel Kant: Groundwork for the Metaphysics of Morals*. Oxford: Oxford University Press.
Blair, B. 1993. *The Logic of Accidental Nuclear War*. Washington, DC: Brookings Institute.
Bostrom, N. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
Boulanin, V. 2020. "*Artificial Intelligence, Strategic Stability and Nuclear Risk*." SIPRI Report, June 2020.
Brodie, B., and B. Fawn. 2010. *From Crossbow to H-Bomb*. Bloomington: Indiana.
Bronfenbrenner, U. 2010. "Mirror Image in Soviet-American Relations: A Social Psychologist's Report." *Journal of Social Issues* 17 (3): 5–56. doi:10.1111/j.1540-4560.1961.tb01682.x.
Buchanan, B., and F. Cunningham. 2020. "Preparing the Cyber Battlefield: Assessing a Novel Escalation Risk in a Sino-American Crisis." *Texas National Security Review* 3 (4): 55–81.
Butterfield, H. 1951. *History and Human Relations*. London: Collins.
Calvin, W. 2006. "The Emergence of Intelligence." *Scientific American* 271 (4): 100–107. doi:10.1038/scientificamerican1094-100.
Clarke, L. 2010. *Worst Cases*. Chicago: University of Chicago Press.
Collins, J., N. Hall, and L. A. Paul, eds. 2004. *Causation and Counterfactuals*. MIT University Press.
Connelly, M., M. Fay, G. Ferrini, M. Kaufman, W. Leonard, H. Monsky, R. Musto, et al. 2012. "'General, I Have Fought Just as Many Nuclear Wars as You Have': Forecasts, Future Scenarios, and the Politics of Armageddon." *American History Review* 117 (5): 1431–1460. DOI:10.1093/ahr/117.5.1431.
Considine, L. 2021. "Narrative and Nuclear Weapons Politics: The Entelechial Force of the Nuclear Origin Myth." *International Theory* 1–20. doi:10.1017/S1752971921000257.
Davis, P., and P. Bracken. 2022. "Artificial Intelligence for Wargaming and Modeling." *The Journal of Defense Modeling and Simulation* 154851292110731. doi:10.1177/15485129211073126.
Dawes, R. 1988. *Rational Choice in an Uncertain World*. San Diageo: Harcourt Brace Jovanovich.
De Dreu, C. 2003. "Time Pressure and Closing of the Mind in Negotiation." *Organizational Behavior and Human Decision Processes* 91 (2): 280–295. doi:10.1016/S0749-5978(03)00022-0.
Favaro, M. 2021. *Weapons of Mass Distortion: A New Approach to Emerging Technologies, Risk Reduction, and the Global Nuclear Order*. London: Center for Science & Security Studies, King's College London.
Fearon, J. 1996. "Causes and Counterfactuals in Social Science: Exploring an Analogy between Cellular Automata and Historical Processes." In *Counterfactual Thought Experiments in World Politics*, edited by P. Tetlock and A. Belkin, 39–69. Princeton: Princeton University Press.
Ferguson, C., and W. Potter. 2004. *The Four Faces of Nuclear Terrorism*. Monterey, CA: Center for Nonproliferation Studies and Nuclear Threat Initiative.
Ferguson, N. 2014. *Virtual History: Alternatives and Counterfactuals*. Penguin.
Fischoff, B. 1975. "Hindsight Is Not Equal to Foresight: The Effect of Outcome Knowledge on Judgment under Uncertainty." *Journal of Experimental Psychology* 1 (2): 288–299.
Fitzpatrick, M. 2009. *The World After: Proliferation, Deterrence and Disarmament if the Nuclear Taboo Is Broken*. Paris: Ifri Security Studies Centre.
Forest, J. 2012. "Framework for Analyzing the Future Threat of WMD Terrorism." *Journal of Strategic Security* 5 (4): 51–68. doi:10.5038/1944-0472.5.4.4.
Francis, G. 2020. "History and the Unanswered Questions of the Nuclear Age." In *The Age of Hiroshima*, edited by J. Ikenberry, 294–311. Princeton: Princeton University Press.
Freedman, L. 1989. *The Evolution of Nuclear Strategy*. 2nd ed. London, UK: Palgrave Macmillan.
Futter, A., and B. Zala. 2021. "Strategic Non-Nuclear Weapons and the Onset of a Third Nuclear Age." *European Journal of International Security* 6 (3): 257–277. doi:10.1017/eis.2021.2.
Gamel, K. 2017. "US Forces Korea Warns of Fake Evacuation Messages." *Stars & Stripes*, 21 September 2017. https://www.stripes.com/theaters/asia_pacific/us-forces-korea-warns-of-fake-evacuation-messages-1.488792

Garfinkel, B., and A. Dafoe. 2019. "Ow Does the offense-defense Balance Scale?" *Journal of Strategic Studies* 42 (6): 736–763. doi:10.1080/01402390.2019.1631810.
George, A., and R. Smoke. 1974. *Deterrence in American Foreign Policy: Theory & Practice*. New York: Columbia University Press.
Goodman, N. 1983. *Fact, Fiction, and Forecast*. Cambridge: Harvard University Press.
Gould, J. D. 1969. "Hypothetical History." *Economic History Review* 22 (2): 199–200. doi:10.2307/2593767.
Green, B., and A. Long. 2019/20. "Conceal or Reveal? Managing Clandestine Military Capabilities in Peacetime Competition." *International Security* 44 (3): 48–83. doi:10.1162/isec_a_00367.
Hawkins, S., and R. Hastie. 1990. "Hindsight: Biased Judgments of past Events after the Outcomes are Known." *Psychological Bulletin* 107 (2): 311–327. doi:10.1037/0033-2909.107.3.311.
Heidegger, M. 1962. *Being and Time*, ed. J. MacQuarrie and E. Robinson. London: SCM Press.
Herz, J. 1951. *Political Realism and Political Idealism: A Study in Theories and Realities*. Chicago: University of Chicago Press.
Horowitz, M., G. C. Allen, E. B. Kania, and P. Scharre. 2018. "Strategic Competition in an Era of Artificial Intelligence." *Artificial Intelligence and International Security, Center for New American Security* July
Hume, D. 1999. *An Enquiry Concerning Human Understanding: A Critical Edition*, edited by T. L. Beauchamp. Oxford: Oxford University Press.
Jacobs, R. 2010. *Filling the Hole in the Nuclear Future: Art and Popular Culture Respond to the Bomb*. New York: Lexington Books.
James, W. 1977. "The Compounding of Consciousness". In *The Writings of William James: A Comprehensive Edition*, edited by, J. J. McDermott. 560. Chicago: University of Chicago Press. Original work published in 1909
Jervis, R. 1976. *Perception and Misperception in International Politics*. Princeton: Princeton University Press.
Jervis, J. 1978. "Cooperation under the Security Dilemma." *World Politics* 30 (2): 169–214. doi:10.2307/2009958.
Jervis, R. 1985a. "Introduction: Approach and Assumptions." In *Psychology and Deterrence*, edited by J. Stein, R. Lebow, and R. Jervis, 1–13. New York: Johns Hopkins University Press .
Jervis, R. 1985b. "Perceiving and Coping with Threat." In *Psychology and Deterrence*, edited by J. Stein, R. Lebow, and R. Jervis, 13–33. New York: Johns Hopkins University Press.
Jervis, R. 1989. "Rational Deterrence: Theory & Evidence." *World Politics* 41 (2): 193–194. doi:10.2307/2010407.
Jia, D., and H. Zhou. 2016. "The Future 20-30 Years Will Initiate Military Transformation". *China Military Online*, June 2.
Johnson, D. 2004. *Overconfidence and War*. Cambridge: Harvard University Press.
Johnson, J. 2019. "Chinese Evolving Approaches to Nuclear "War-Fighting": An Emerging Intense US-China Security Dilemma and Threats to Crisis Stability in the Asia Pacific." *Asian Security* 15 (3): 215–232. doi:10.1080/14799855.2018.1443915.
Johnson, J., and E. Krabill. 2020. "AI, Cyberspace, and Nuclear Weapons." *War on the Rocks*, January.
Johnson, J. 2020. "Artificial Intelligence in Nuclear Warfare: A Perfect Storm of Instability?" *The Washington Quarterly* 43 (2): 197–211. doi:10.1080/0163660X.2020.1770968.
Johnson, J. 2021a. "'Catalytic Nuclear War' in the Age of Artificial Intelligence & Autonomy: Emerging Military Technology and Escalation Risk between nuclear-armed States." *Journal of Strategic Studies* 1–41. Online first. doi:10.1080/01402390.2020.1867541.
Johnson, J. 2021b. "Inadvertent Escalation in the Age of Intelligence Machines: A New Model for Nuclear Risk in the Digital Age." *European Journal of International Security* Online first. doi:10.1017/eis.2021.23.
Kahn, H. 1960. *On Thermonuclear War*. New Brunswick, NJ: Princeton University Press.
Kahneman, D. 1995. "Varieties in Counterfactual Thinking." In *What Might Have Been: The Social Psychology of Counterfactual Thinking*, edited by N. J. Roese and J. M. Olson, 375–396. Mahwah, NJ: Erlbaum.

Kaplan, M., T. Wanshula, and M. Zanna. 1993. "Time Pressure and Information Integration in Social Judgment." In *Time Pressure and Stress in Human Judgment and Decision Making*, edited by O. Svenson and J. Maule, 255–267. Boston, MA: Springer.
Kershaw, I. 2007. *Fateful Choices: Ten Decisions that Changed the World, 1940–1941*. New York: Penguin.
King, G., R. Keohane, and S. Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
Kissinger, H. 1993. *Diplomacy*. New York, NY: Simon and Schuster.
Knight, F. 1921. *Risk, Uncertainty and Profit*. New York: Dover Publications.
Knopf, J. 2012. "The Concept of Nuclear Learning." *Nonproliferation Review* 19 (1): 79–93. doi:10.1080/10736700.2012.655088.
Kobe, D. 1962. "A Theory of Catalytic War." *The Journal of Conflict Resolution* 6 (2): 443–457. doi:10.1177/002200276200600203.
Kroenig, M. 2021. "Will Emerging Technology Cause Nuclear War? Bringing Geopolitics Back." *Strategic Studies Quarterly* 15 (4): 59–73.
Kuhn, T. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
Lebow, R., and J. G. Stein. 1989. "Rational Deterrence Theory: I Think Therefore I Deter." *World Politics* 41 (2): 208–224. doi:10.2307/2010408.
Lebow, R., and J. Stein. 1996. "Back to the Past: Counterfactuals and the Cuban Missile Crisis." In *Counterfactual Thought Experiments in World Politics*, edited by P. Tetlock and A. Belkin, 119–149. Princeton: Princeton University Press.
Lebow, R. 2010. *Forbidden Fruit: Counterfactuals & International Relations*. Princeton: Princeton University Press.
Lebow, R. 2015. "Counterfactuals and Security Studies." *Security Studies* 24 (3): 403–441. doi:10.1080/09636412.2015.1070605.
Lebow, R., and B. Pelopidas. Forthcoming. "Facing Nuclear War: Luck, Learning, and the Cuban Missile Crisis." In *Oxford Handbook of History and International Relations*, edited by B. Reus-Smit, et al. Oxford: Oxford University Press.
Levy, J. 2015. "Counterfactuals, Causal Inference, and Historical Analysis." *Security Studies* 24 (3): 378–402. doi:10.1080/09636412.2015.1070602.
Lewis, P., H. Williams, B. Pelopidas, and S. Aghlani. 2014. *Too Close for Comfort: Cases of near Nuclear Use and Options for Policy*. London: Chatham House.
Lieber, K. 2008. *War, and the Engineers: The Primacy of Politics over Technology*. Ithaca, NY: Cornell University Press.
Lin, H. 2012. "Escalation Dynamics and Conflict Termination in Cyberspace." *Strategic Studies Quarterly* 6 (3): 46–70.
Lin, H. 2021. *Cyber Threats and Nuclear Weapons*. Stanford, CA: Hoover Institution Press.
McCloskey, D. 1990. "History, Differential Equations, and the Problem of Narration." *History and Theory* 30 (1): 21–36. doi:10.2307/2505289.
McDonald, J., and M. Bell. 2019. "How to Think about Nuclear Crises?" *Texas National Security Review* 2 (2): 41–65.
Mearsheimer, J. 1990. "Back to the Future: Instability in Europe after the Cold War." *International Security* 15 (1): 5–56. doi:10.2307/2538981.
Nisbett, R., and L. Ross. 1980. *Human Inference: Strategies and Shortcomings of Social Judgment*. New Jersey: Prentice-Hill.
Pelopidas, B. 2016. "Nuclear Weapons Scholarship as a Case of Self-Censorship in Security Studies." *Journal of Global Security Studies* 1 (4): 326–336. doi:10.1093/jogss/ogw017.
Pelopidas, B. 2017. "The Unbearable Lightness of Luck. Three Sources of Overconfidence in the Controllability of Nuclear Crises." *European Journal of International Security* 2 (2): 240–262. doi:10.1017/eis.2017.6.
Pelopidas, B. 2020. "Power, Luck, and Scholarly Responsibility at the End of the World(s)." *International Theory* 12 (3): 459–470. doi:10.1017/S1752971920000299.
Perrow, C. 1984. *Normal Accidents: Living with High-Risk Technologies*. Princeton: Princeton University Press.

Popper, K. 2005. *The Logic if Scientific Discovery*. London: Routledge.
Rodd, D. 1977. "The Intuitive Psychologist and His Shortcomings: Distortions in Attribution Process." *Advances in Experimental Social Psychology* 10: 173–220.
Rosen, S. 2010. "The Impact of the Office of Net Assessment on the American Military in the Matter of the Revolution in Military Affairs." *Journal of Strategic Studies* 33 (4): 469–482. doi:10.1080/01402390.2010.489704.
Rosenblatt, A., J. Greenberg, S. Solomon, T. Pyszcynski, M. V. Veeder, S. Kirkland, and D. Lyon. 1990. "Evidence for Terror Management II: The Effects of Mortality Salience on Reactions to Those Who Threaten or Bolster the Cultural World View." *Journal of Personality and Social Psychology* 58 (2): 308–318. doi:10.1037/0022-3514.58.2.308.
Sagan, S. 1993. *The Limits of Safety. Organizations, Accidents and Nuclear Weapons*. Princeton: Princeton University Press.
Schelling, T. 1960. *The Strategy of Conflict*. Cambridge: Harvard University Press.
Schneider, F., and J. Sherman. 2021. "Bases for Trust in a Supply Chain." *Lawfare*, 1 February 2021. https://www.lawfareblog.com/bases-trust-supply-chain
Schoemaker, P. 1991. "When and How to Use Scenario Planning: A Heuristic Approach with Illustration." *Journal of Forecasting* 10 (6): 549–564. doi:10.1002/for.3980100602.
Schwartz, P. 1991. *The Art of the Long View*. New York: Doubleday/Currency.
Shao, C., G. L. Ciampaglia, O. Varol, K.-C. Yang, A. Flammini, F. Menczer. 2018. "The Spread of low-credibility Content by Social Bots." *Nature Communications* 9 (4787): 1–8. DOI:10.1038/s41467-018-06930-7.
Sherwood-Randall, E. 2020. "The Age of Strategic Instability: How Novel Technologies Disrupt the Nuclear Balance." *Snapshot, Foreign Affairs*, July 21.
Tannenwald, N. 2018. "How Strong Is the Nuclear Taboo Today?" *The Washington Quarterly* 41 (3): 89–109. doi:10.1080/0163660X.2018.1520553.
Tetlock, P. 1992. "Good Judgment in International Politics: Three Psychological Perspectives." *Political Psychology* 13 (3): 517–539. doi:10.2307/3791611.
Tetlock, P., and A. Belkin, eds. 1996. *Counterfactual Thought Experiments in World Politics*. Princeton, NJ: Princeton University Press.
Tetlock, P. 1998. "Close-Call Counterfactuals & Belief System Defenses: I Was Not Almost, but I Was Almost Right." *Journal of Personality and Social Psychology* 75 (3): 230–242. doi:10.1037/0022-3514.75.3.639.
Tetlock, P. 2005. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton: Princeton University Press.
Trinkunas, H., H. Lin, and B. Loehrke. 2020. *Three Tweets to Midnight: Effects of the Global Information Ecosystem on the Risk of Nuclear Conflict*. Stanford, CA: Hoover Institution Press.
Trivers, R. 2014. *The Folly of Fools: The Logic of Deceit and Self-Deception in Human Life*. New York: Basic Books.
Tversky, A., and D. Kahneman. 1974. "Judgement under Uncertainty: Heuristics and Biases." *Science* 185 (27): 1124–1131. doi:10.1126/science.185.4157.1124.
Vacher, J., A. Meso, L. Perrinet, and G. Peyré. 2018. "Bayesian Modeling of Motion Perception Using Dynamical Stochastic Textures." *Neural Computation* 30 (12): 3355–3392. doi:10.1162/neco_a_01142.
Waltz, K. 1979. *Theory of International Politics*. New York: Random House.
Weber, S. 1996. "Counterfactuals, Past, and Future." In *Counterfactual Thought Experiments in World Politics*, edited by P. Tetlock and A. Belkin, 268–291. Princeton: Princeton University Press.
Wohlstetter, R. 1962. *Pearl Harbor*. Stanford: Stanford University Press.
Zuboff, S. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs.
Zwald, Z. 2013. "Imaginary Nuclear Conflicts: Explaining Deterrence Policy Preference Formation". *Security Studies* 22 (4): 640–671. doi:10.1080/09636412.2013.844519.