

## Poking Counterfactual Holes in Covering Laws: Cognitive Styles and Historical Reasoning

PHILIP E. TETLOCK and RICHARD NED LEBOW *The Ohio State University*

**W**e report a series of studies of historical reasoning among professional observers of world politics. The correlational studies demonstrate that experts with strong theoretical commitments to a covering law and cognitive-stylistic preferences for explanatory closure are more likely to reject close-call counterfactuals that imply that "already explained" historical outcomes could easily have taken radically different forms. The experimental studies suggest that counterfactual reasoning is not totally theory-driven: Many experts are capable of surprising themselves when encouraged to imagine the implications of particular what-if scenarios. Yet, there is a downside to openness to historical contingency. The more effort experts allocate to exploring counterfactual worlds, the greater is the risk that they will assign too much subjective probability to too many scenarios. We close by defining good judgment as a reflective-equilibrium process of balancing the conflicting causal intuitions primed by complementary factual and counterfactual posings of historical questions.

**T**he philosopher of science Carl Hempel (1965) advanced the controversial thesis that history, properly understood, is a prime candidate for reduction to the laws of social science. Events can only be considered "explained" when they have been assimilated into syllogisms that meet three conditions: (1) causally efficacious covering laws serve as major premises, (2) carefully abstracted antecedent conditions that prevail at given times and places serve as minor premises, and (3) the conclusions follow as the inexorable or at least likely result of the hypothesized laws operating on the specified antecedents. For better or for worse, and many historians believe the latter (Dray 1989), this covering-law model captures the approach of many social scientists to the explanation of historical trends and patterns (Goldstone n.d.; Mokyr n.d.). This kind of disciplinary tension is familiar: between idiosyncratic and nomothetic camps, between particularizers who complain that theorists neglect critical complexities in their eagerness to assimilate history into their favorite explanatory templates, and generalizers who complain that particularizers are so immersed in idiosyncratic detail that they miss the big theoretical picture.

This article explores a recurring source of disagreement between generalizers and particularizers: the soundness of close-call counterfactual scenarios that imply that, with only minimal rewriting of antecedent conditions, history could have been rerouted down different, sometimes radically different, event paths. Close-call counterfactuals are often focal points of disagreement for two reasons.

First, Fogel (1964) and Fearon (1991), among others, demonstrate that all causal inference from history

ultimately rests on counterfactual claims about what would or might or could have happened in hypothetical worlds to which scholars have no direct empirical access. This is not to say that evidence is irrelevant. A variety of empirical and logical criteria can be used to differentiate more from less compelling counterfactual claims (see the essays in Tetlock and Belkin 1996). But disputes over the relative soundness of competing counterfactuals are often notoriously resistant to consensual resolution. Examples include the seemingly interminable debates over the avertability or inevitability of both sweeping historical transformations, such as the European rise to global hegemony (Landes 1997; Mokyr n.d.), and highly specific events, such as the onset of World War I or the end-game moves of the Cold War (Lebow 2001).

Second, close-call counterfactuals assign critical roles to causes that covering-law theorists tend to disparage as trivial, impossible to categorize, and best relegated to error variance (see McCloskey, 1991). The quest for parsimonious laws governing political processes becomes progressively more hopeless to the degree that the triumph of civilizations hinges on horseshoe-nail (or butterfly) effects in battles; the outbreak of world wars hinges on royal carriage drivers making a wrong turn in downtown Sarajevo or an emotionally unstable young man gaining admission to a Viennese art school; or the outcome of the Cuban missile crisis hinges on the interpersonal chemistry among presidential advisers.

We work from the uncontroversial premise that there is an ineradicable element of subjectivity in these debates over historical causality. Accordingly, a strong case can be made for rigorously scrutinizing potential cognitive biases in how historical observers go about evaluating the relative plausibility of competing what-if scenarios. We report two sets of empirical studies, one correlational and the other experimental. The correlational research explores the degree to which professional observers of world politics rely on abstract covering laws in assessing what was possible at specific times and places. These studies demonstrate that, across an array of contexts, the more credence observ-

---

Philip E. Tetlock is Burt Professor of Psychology and Political Science, and Richard Ned Lebow is Professor of Political Science, History, and Psychology, the Mershon Center, The Ohio State University, Columbus, OH 43201.

This research was supported by the Mershon Center, The Ohio State University, as well as by the MacArthur Foundation. We express our deep gratitude to the many colleagues who took the time to respond to our questions. We also appreciate the helpful advice of Geoffrey Parker, Daniel Kahneman, the anonymous reviewers, and Ada Finifter.

ers place in relevant covering laws, and the stronger their cognitive-stylistic preferences for explanatory closure, the more likely they are to be guided by those covering laws in judging what could have been, and the less likely they are to make judgments of historical contingency on a case-by-case basis.

Whereas the correlational studies rely on naturally occurring variation in theoretical beliefs and cognitive-stylistic preferences to illustrate the deductive, top-down character of counterfactual reasoning and the power of covering-law preconceptions to shape the conclusions that experts draw about what was historically possible, the experimental studies encourage experts to perform, in effect, an unnatural cognitive act: to give more thought than they normally would to alternative paths that history could have followed. The goal is to test the notion that counterfactual reasoning also can take an inductive, bottom-up form. In other words, the mental processes of imagining specific counterfactual scenarios can induce us to change our mind and become more circumspect about the power of our favorite causal generalizations to delimit the range of historical possibilities. Experiment 1 asks foreign policy experts to consider various scenarios, supported by varying amounts of detail, that "undo" the peaceful resolution of the Cuban missile crisis. The net effect was to increase their perceptions of the potential for alternative, more violent, endings. Experiment 2 shows that the manipulations need not be heavy handed and that no new information need even be presented. Experts were asked to draw on their own historical knowledge in searching for possible pathways to more violent endings of the crisis, and then to unpack these possibilities into progressively more differentiated subsets; the net effect was inflation of the subjective probabilities of those alternative outcomes. Moreover, consistent with Tversky and Fox's (1995) support theory, the more extensive the unpacking, the greater was the resulting inflation of subjective probabilities. Experiment 3 demonstrates that these effects are not peculiar to the relatively recent and brief Cuban missile crisis. It provides a conceptual replication in the context of the debate about the rise of the West, the historical transformation over several centuries of Western Europe from cultural backwater to global hegemon.

Taken together, the results give real-world empirical content to a cognitive account of the construction of historical knowledge that specifies the mental processes by which observers apply causal generalizations to history and judge the acceptability of close-call counterfactuals that circumscribe those generalizations. Historical observers may indeed often be prisoners of their preconceptions (Tetlock 1999), but they hold the keys to their own release. They are not constrained to explain only what happened. We are endowed with the imaginative capacity to envision alternative paths that history could have taken. The more energy we direct to elaborating dissonant alternatives, the weaker is the hold of our preconceptions (judging by the dwindling regression coefficients) on our judgments of contingency and possibility. But this

liberation has a price. The more effort we devote to unpacking counterfactual scenarios, the more contradictory are our resulting judgments of history. Our findings suggest that historical observers of world politics confront an inescapable trade-off between being closed-minded but logically coherent or open-minded but logically incoherent.

## THE CORRELATIONAL STUDIES: APPLYING ABSTRACT COVERING LAWS TO SPECIFIC HISTORICAL CASES

### Guiding Theory

From a neopositivist perspective on historical explanation (see Hempel 1965), theorists are under absolutely no logical obligation to apply covering laws to any specific event. Covering laws in social science are best viewed both as probabilistic and as applicable to sets of events, not individual occurrences. Logical necessity to the side, however, there may well be considerable psychological pressure on observers to achieve closure in their historical explanations.

As Herbert Simon (1996) and colleagues have shown, even highly trained experts function like boundedly rational satisficers who try to keep the number of variables and amount of information that must be integrated to a reasonable minimum. Also, a growing psychological literature on cognitive style reveals marked individual differences in the motivation to achieve simplicity and closure in the characterization and explanation of events (Kruglanski and Webster 1996; Suedfeld and Tetlock 2001). Experimental and field research finds, for example, that people who score high on self-report measures of the need for closure tend to rely on prior beliefs in solving unfamiliar problems and in evaluating dissonant arguments, prefer parsimonious interpretations of evidence that invoke as few causal constructs as possible, and prefer deterministic accounts that downplay probabilistic qualifiers. This body of work suggests that professional observers of world politics with a strong need for closure should find close-call counterfactuals vexing to the degree these imply that minor tinkering with antecedent conditions could have undone major outcomes that the observers felt they had already comfortably assimilated into favorite covering-law schemes.

Exactly how do experts resist theoretically subversive counterfactuals? From a strictly logical perspective, there are three possible lines of defense (Lewis 1973). From a psychological perspective, especially one grounded in cognitive consistency theory, there are good reasons for expecting experts to mobilize all three whenever experts feel core theoretical commitments are under siege (see Jervis 1976 on belief-system overkill). The lines of defense are as follows.

First, challenge the mutability of the antecedent. For example, insist that it is hard to imagine Kennedy heeding his more hawkish advisers during the Cuban missile crisis. This defense often seems contrived, however, even to those who value closure. All but the most implacably La Placean determinists recognize

that it is exceedingly implausible to insist that each and every microscopic detail surrounding key events had to take the form it did.

Second, challenge the connecting principles that bridge antecedent and consequent. Counterfactual arguments are inherently elliptical; it is impractical to spell out all the background assumptions that must be satisfied to sustain even the simplest antecedent-consequent linkages. Consider the claim that if Kennedy had heeded his hawkish advisers during the missile crisis, and attacked Soviet bases in Cuba, World War III would have ensued. Deterrence theorists can counter that war was not inevitable because the Soviets recognized their strategic inferiority vis-à-vis the United States. They would have grudgingly acquiesced to the *fait accompli* and forbidden their ground forces in Cuba to use nuclear-tipped Luna missiles against the American armada.

Third, concede that the antecedents may be mutable and that connecting principles may be sound, but still insist that history would have been thrown off track only temporarily, and the proposed changes would have had little long-term significance. This defense is designated a second-order counterfactual because it undoes the undoing of history implied by the original scenario. For example, even if Genghis Khan had lived and the Mongols had mauled Western Europe, the resilient Europeans would have bounced back quickly.

## Hypothesis

The preceding argument lays the conceptual groundwork for the two hypotheses tested in each of the three correlational studies. The covering-law hypothesis predicts a main effect: The more confidence an expert places in any given causal generalization, the more likely s/he is to invoke all three lines of belief-system defense against close-call scenarios that undercut the applicability of the generalization to historical outcomes previously thought covered by that generalization. The cognitive-style hypothesis predicts an interactive effect: Experts with a strong need for explanatory closure are likely to invoke all three defenses against dissonant scenarios.

## Research Design and Logic of Analysis

The correlational studies probe reactions to close-call scenarios that undercut three categories of covering law. Study 1 targets variants of nuclear deterrence theory that stress the power of mutual assured destruction to induce even minimally rational actors to act with self-restraint (see, e.g., the instructive exchange in Sagan and Waltz 1995). Scenarios that imply we were on the brink of nuclear war at numerous junctures in the Cold War should seem far-fetched to theorists who posit the robustness of nuclear deterrence. Looking back on the Cold War, these theorists should find it hard to imagine that either superpower ever would have allowed a crisis to escalate into a nuclear war (just as, looking forward, they find it hard to worry about the dangers of nuclear proliferation).

Study 2 targets variants of the neorealist balancing hypothesis. That is, when one state threatens to become too powerful and to dominate the entire international system, other states—rational, self-preserving actors as they are posited to be—coalesce against it, thereby preserving the balance of power (cf. Vasquez 1997). From this standpoint, the failure of aspiring hegemony, such as Philip II, Napoleon, or Hitler, is predetermined by a fundamental law of world politics, and close-call claims that these bids to dominate Europe might have succeeded are misinformed.

Study 3 targets a pair of covering laws that loom large in debates on the rise and fall of civilizations. The first is that history, in the long term, is an efficient process of winnowing out maladaptive forms of social organization. The second is that Western societies, with their emphasis on the rule of law, property rights, free markets, and the practical applications of science, are better adapted to prevail in long-term competition with other civilizations (Mokyr n.d.). It is no accident that, between 1500 and 1800, Europeans came to wield global influence so disproportionate to their numbers. The rise of the West, at the expense of the “Rest,” reveals the fundamental superiority of Western values and institutions. Close-call counterfactuals that purport to undo the rise of the West by minor mental mutations of antecedent conditions—more aggressive Islamic incursions into France in the eighth century or less isolationist Chinese emperors in the fifteenth century—are merely whimsical excursions into what-if history.

## Methods and Measures

Participants were drawn from overlapping populations of scholars who specialize in diplomatic and military history, security studies, and international relations. The 87 participants in correlational studies 1 and 2, which focus on neorealist balancing and nuclear deterrence, were randomly sampled from Division 18 (International Conflict) and Division 19 (International Security and Arms Control) of the American Political Science Association and from the Society of Diplomatic and Military Historians. The 63 participants for study 3, which focuses on the rise of the West, were randomly drawn from the membership roster of the World History Association. All respondents were contacted by either postal mail or e-mail, and they were promised both anonymity and detailed feedback on the purposes of the study. The response rate for studies 1 and 2 was 29%; for study 3, 31%. Assessing selection bias as a result of the relatively low response rates was problematic because of the limited information available on participating and nonparticipating individuals in our sample. Available results suggest, however, only slight overrepresentation of academics employed in large research universities.

**Covering-Law Beliefs.** These measures assessed endorsement of the covering laws on nine-point agree-disagree scales. For neorealist balancing there were four items. (a) “For all the talk about a new world



order, world politics is still essentially anarchic—the strong do what they will and the weak accept what they must”; (b) “Whenever one state starts to become too powerful, other states find a way of combining forces and preventing it from dominating them”; (c) “The security policies of states are often driven by morality, not just by rational calculations of the impact of those policies on the balance of power”; (d) “It is naive to suppose that the failure of would-be conquerors such as Philip II, Napoleon, and Hitler to achieve lasting dominance in Europe was predetermined by balance-of-power politics—it might just have been an accident.”

For nuclear deterrence there were three items. (a) “For all the talk about the risk of nuclear accidents, the U.S.A. and U.S.S.R. never really came close to nuclear war”; (b) “Nuclear weapons played a key role in moderating the behavior of both the American and Soviet governments during the Cold War”; (c) “It is unrealistic to assume that leaders working under great stress will always act with great restraint in crises that raise the risk of the use of nuclear weapons.”

Two items dealt with the survival of the fittest civilizations. (a) “History is, in the long run, an efficient process of winnowing out maladaptive forms of social organization”; (b) “Western societies and institutions, with their greater emphasis on the rule of law, property rights, free markets, and the practical applications of science, were better adapted to prevail in long-term competition with other civilizations.”

**The Need for Explanatory Closure.** Respondents gave answers on a nine-point agree-disagree scale to nine items: (a) “I think that having clear rules and order at work is essential for success”; (b) “Even after I have made up my mind about something, I am always eager to consider a different opinion”; (c) “I dislike questions that can be answered in many different ways”; (d) “I usually make important decisions quickly and confidently”; (e) “When considering most conflict situations, I can usually see how both sides could be right”; (f) “I prefer interacting with people whose opinions are very different from my own”; (g) “When trying to solve a problem I often see so many possible options that it is confusing”; (h) “Scholars are usually at greater risk of exaggerating the complexity of political processes than they are of underestimating the complexity of those processes”; (i) “Isaiah Berlin [1997] classified intellectuals as hedgehogs or foxes. A hedgehog knows one big thing and tries to integrate the diversity of the world into a comprehensive and parsimonious vision whereas a fox knows many small things and tries to improvise explanations on a case-by-case basis. I would place myself toward the hedgehog or fox style of thinking about politics.”

**Beliefs about Close-Call Counterfactuals.** The first set of scenarios undid the outcomes of either World War I or II: (a) “If Germany had proceeded with its invasion of France on August 2, 1914, but had respected the Belgian neutrality, Britain would not have entered the war, and France would have quickly fallen”; (b) “If the German High Command had implemented the Schlieffen Plan more aggressively in August, 1914, the miracle

of the Marne would have been impossible, and Paris would have fallen”; (c) “If Germany had avoided antagonizing the United States by meddling in Mexico and by initiating unrestricted submarine warfare, the United States would not have entered World War I and Germany would have prevailed against the French and British in its spring offensive of 1918”; (d) “If Hitler had not invaded the Soviet Union and concentrated German resources on defeating the British, Germany would have defeated Britain”; (e) “If Hitler had more consistently focused on taking Moscow in the summer of 1941, he could have knocked the Soviet Union out of the war”; (f) “If Hitler had not declared war on the United States on December 11, 1941, the British and the Soviets could never have defeated Nazi Germany.”

The second set of counterfactuals explored the feasibility of the Cold War becoming thermonuclear: (a) “If Stalin had lived several years longer (surviving his stroke but in an irrational state of mind that encouraged high-risk adventures), World War III could easily have broken out in the mid-1950’s”; (b) “If bad weather had delayed the discovery by U-2 reconnaissance planes of Soviet missiles in Cuba until the missiles were operational, the Soviets would have refused American demands to dismantle and withdraw the weapons”; (c) “If the Soviets had refused to withdraw their missiles, the U.S. would have launched air strikes against the Soviet bases”; (d) “If the U.S. had launched such air strikes, the Soviet commanders in Cuba would have launched at least some missiles at the eastern seaboard of the United States”; (e) “If the Soviets had fired Cuban-based nuclear missiles at American cities, retaliatory nuclear strikes would have been launched at Soviet cities”; (f) “If Soviet hardliners had taken charge of the Communist Party in the mid-1980s, the Cold War—far from ending peacefully and quickly—would have intensified.”

The third set of counterfactuals explored the feasibility of unmaking the West through hypothetical interventions that either enfeebled Europe or empowered rival civilizations: (a) “If China had had, at key junctures, emperors more sympathetic to economic and technological development, it could have emerged as the world’s first superpower”; (b) “If the Mongols had continued their advance into central and western Europe and not been distracted by the death of Genghis Khan, later European development would have been impossible”; (c) “If Islamic armies had made a serious attempt to conquer France and Italy in the 8th century, later European development could have been radically side-tracked”; (d) “If the Black Death had been even more lethal, killing, say, 70% of the population, Europe could not have arisen as the dominant region in the second half of the millennium.”

For each scenario, experts were asked to make three judgments on nine-point scales. (1) “How plausible was the antecedent condition of the argument? (Do we have to ‘re-write’ a little or a lot of history?)” (2) “Assuming the plausibility of the antecedent, how likely was the hypothesized consequence?” (3) “Assuming the plausibility of the hypothesized conse-

**TABLE 1. Predicting Resistance to Close-Call Counterfactuals**

Covering Law	<i>b</i>	SE	<i>t</i>
Neorealist balancing	0.96***	0.30	3.18
Cognitive style	0.35	0.29	1.20
Balancing × style	0.74**	0.36	2.07
<i>n</i> = 87			
<i>R</i> <sup>2</sup> = 0.47			
Nuclear deterrence	0.89**	0.34	2.65
Cognitive style	0.33	0.31	1.07
Deterrence × style	0.69**	0.33	2.06
<i>n</i> = 86			
<i>R</i> <sup>2</sup> = 0.43			
Adaptive advantage of West	0.82**	0.36	2.27
Cognitive style	0.23	0.28	0.83
West × style	0.73*	0.36	2.10
<i>n</i> = 63			
<i>R</i> <sup>2</sup> = 0.41			

Note: The table presents the full ordinary-least-squares results for each of three separate multiple regressions. Each treats resistance to close-call counterfactuals as the dependent variable; independent variables are commitment to a particular school of thought (neorealist balancing, nuclear deterrence, and adaptive advantages of West), cognitive style (need for closure), and a cross-product term designed to capture degree to which resistance is greatest when both theoretical commitment and need for closure are high. The nuclear deterrence and neorealist balancing scales were correlated quite highly ( $r = .57$ ). \* $p \leq .05$ , \*\* $p \leq .01$ , \*\*\* $p \leq .001$ .

quence, what would the long-term ramifications have been?"

## Findings

Correlational analyses revealed that experts invoked all three belief-system defenses against dissonant close-calls but preferred certain defenses. Two of them—challenge the logic of the connecting principles and generate second-order counterfactuals that put history back on track—were widely employed and were tightly linked to the respondent's abstract orientation toward world politics (average  $r = 0.54$  with the covering-law scales). The third defense—challenge the mutability of historical antecedents—was markedly less linked to abstract political orientation (average  $r = 0.29$ ). There is no compelling reason one's theoretical position on the robustness of nuclear deterrence should predict whether one believes Stalin could have survived the cerebral hemorrhage of March 1953 or whether Cuba could have been cloudier on an October day in 1962. The plausibility of most antecedents hinges on specific facts tied to particular times, places, and events; if anything, it is disconcerting that abstract orientation predicts so much variance—8% to 10%—in judgments of the mutability of antecedents.

Table 1 summarizes ordinary-least-squares multiple regression results for a composite dependent variable that additively combines the two most correlated resistance strategies, challenge connecting principles and

invoke second-order counterfactuals. These regressions consistently show that the more committed the scholars were to a school of thought, the more dismissive they were of counterfactuals that undercut historical applications of favorite explanatory generalizations. The more experts believed in the reliability of the balancing hypothesis, the more they rejected the scenarios that implied Germany, with slightly sounder decision making, could have emerged victorious in either of the two world wars and achieved continental hegemony. The more experts believed in the robustness of nuclear deterrence, the more dismissive they were of close-calls that implied the United States and the Soviet Union easily could have slipped into nuclear conflict during the Cold War. The more experts believed in the superiority of Western institutions, the less use they had for counterfactuals that implied other civilizations, with minor twists of fate, could have been contenders for geopolitical dominance.

Table 1 also shows that, as predicted, the interaction between theoretical belief and covering law emerged in all three domains of application: Experts who most valued explanatory closure and parsimony mounted the stiffest resistance to dissonant scenarios. The interaction cannot, moreover, be attributed to a differential restriction of range. Hartley's  $F_{\max}$  test for heterogeneity of variance revealed only minimal differences in the variance of each theoretical belief scale among low versus high scorers on the cognitive-style scales (median split) and equally negligible differences in variation on the cognitive-style scale between low and high scorers on the three theoretical belief scales (no  $p$  value fell below .20). Differential reliability of measures also was not a factor. Cronbach alphas for the three theoretical belief scales were all in the relatively narrow range of 0.70 to 0.80 for both low and high scorers on the cognitive-style scale.

## Discussion

Respondents with a high need for closure were not content just to claim that they had the right explanations of the past. They insisted that they "were not almost wrong," and that the historical outcomes they felt they had explained either were difficult to "undo" or were not undoable for long. As soon as one causal pathway to the theoretically predicted outcome was cut off, another pathway arose, hydra-like, in a second-order counterfactual (Tetlock 1998).

These findings shed new light on some old observations: It is easy to extract lessons from history that reinforce our ideological or theoretical stance (Jervis 1976; Neustadt and May 1986), and it is difficult to avoid the temptation of selection bias and theoretically self-serving reasoning in defense of that stance (Lustick 1996). The data also put in new psychological perspective some old epistemological and methodological controversies. The patterns of counterfactual inference documented here can be viewed as integral parts of the cognitive mechanisms that produce selection bias in the use of historiography. Theoretically committed observers feel justified in giving short shrift to historical

accounts that dwell on inconsequential contingencies and frivolous what-if scenarios.

The cognitive patterns documented here also encourage the conceptual stretching of theories that some scholars identify as signs of a degenerative research program (Vasquez 1997). Stretching can be facilitated either by dismissing dissonant scenarios ("my theory already explains  $x$ , so do not bother me with counterfactuals that imply I was almost wrong") or by embracing consonant close-calls ("my theory is admittedly hard-pressed to explain  $x$ , but  $\sim x$  predicted by my theory almost occurred, so my theory should get some credit"). These arguments raise a fundamental challenge to the discipline: How can we avoid becoming prisoners of our preconceptions, trapped in a self-serving cycle of filling in the missing counterfactual control conditions of history with theory-scripted scenarios, and then deploying that same theory-driven content, in other contexts, to justify the theory itself?

## EXPERIMENTAL STUDIES: THE POWER OF COUNTERFACTUALS TO CHANGE MINDS

### Experiment 1: The Cuban Missile Crisis

**Guiding Theory.** Although experts tend to dismiss dissonant close-call counterfactuals, hefty regression coefficients do not preclude the possibility that these scenarios have some measurable effect on their judgment of historical contingency. Indeed, the psychological literature suggests that causality operates in precisely this direction. Laboratory experiments repeatedly find support for the prediction—derived from Tversky and Kahneman's (1983) classic work on the availability and simulation heuristics—that merely imagining multievent scenarios increases the perceived likelihood of the component events (Koehler 1991). Scenarios tend to be mentally sticky: Once they have been run through our "mental software," they leave psychological traces in the form of images, thoughts, and emotions that can shape a host of subsequent causal-propensity and subjective-probability judgments.

The literature on heuristics and biases also warns us to expect systematic logical inconsistencies when people judge complex event sequences that require integrating two or more probabilistic linkages. The textbook illustration is the conjunction fallacy (Tversky and Kahneman 1983). Imagine that one randomly constituted group is asked to judge the likelihood of a plausible conjunction of events, such as an earthquake that ruptures a dam, which in turn causes a flood that kills more than 500 people in California. Imagine that another randomly constituted group is asked to judge the likelihood of a flood (produced by any cause) killing more than 500 people in California. The likelihood judgments of the former group typically will exceed those of the latter group by a substantial margin, even though the former group is judging a subset of the class of outcomes being judged by the latter group.

Building on this work, Tversky and Fox's (1995) support theory predicts systematic violations of the

extensionality principle of probability theory. The principle requires that if two sets of events have identical coverage, then they must have identical probabilities. Psycho-logic trumps logic here because most people can mobilize mental support more readily for highly specific possibilities than they can for the abstract sets that subsume these possibilities. As a result, people often judge the likelihood of an entire set of possibilities, such as a specific team from a given league winning the championship, to be substantially less likely than the sum of the likelihood values of that set's exclusive and exhaustive components (the probabilities of losses for individual teams that make up the league). In effect, people judge the whole to be less than the sum of its parts and give quite different answers to logically equivalent versions of the same question.

**Hypotheses.** Drawing on the literature on heuristics and biases as well as the work on cognitive styles, we designed Experiment 1 to test two hypotheses. First, thinking about counterfactual scenarios (that pass some minimum plausibility threshold) should tend, on average, to increase the perception that those scenarios once had the potential to materialize and may even once have been more likely than the concatenation of events that actually materialized. Linking this prediction to research on cognitive style, we also expect that the effect should be more pronounced among respondents with low need for closure.

Second, Tetlock (n.d.) shows that there are two logically but not psychologically equivalent methods for scaling experts' perceptions of historical contingency. One imposes a factual framing on the historical question and solicits inevitability-curve judgments. For example, in Experiment 1, experts on the Cuban missile crisis were asked at what point some form of peaceful resolution became inevitable. They then were asked to trace how the subjective probability of that class of outcomes waxed or waned in the preceding days. The other method imposes a "counterfactual" framing on the historical question and solicits impossibility-curve judgments. In Experiment 1, for example, experts also were asked at what point they believe all alternative, more violent endings of the crisis became impossible and then were asked to trace how the subjective likelihood of that class of outcomes waxed or waned in the preceding days.

It was not expected that experts would be blatantly inconsistent: Their judgments of the retrospective likelihood of some form of peaceful outcome between October 16 and 29, 1962, should generally mirror their judgments of the retrospective likelihood of alternative, more violent, outcomes when those judgments are obtained back to back from the same respondents. Logic and psycho-logic should coincide when the principle or binary complementarity is transparently at stake, and experts can plainly see that they are assigning so much probability to both  $x$  and its complement that the sum will exceed 1.0. But logic and psycho-logic do not always coincide. Factual framings of historical questions effectively invite experts to engage in hypothesis-confirming searches for potent causal candidates



that create an inexorable historical momentum toward outcome  $x$ . Analysts feel that they have answered the question when they have convinced themselves that  $x$  had to happen approximately when and in the manner it did.

By contrast, counterfactual framing of historical questions effectively invites analysts to look long and hard for causal candidates that have the potential to reroute events down radically different event paths. Accordingly, we expect systematic anomalies in retrospective likelihood judgments when we compare the judgments of two groups of experts, one of which completed the inevitability curve exercise and the other of which completed the logically redundant impossibility curve exercise, but neither of which had yet seen or worked through the other group's exercise.

We made two "anomaly" predictions. First, systematic violations of binary complementarity should arise at pre-inevitability and pre-impossibility dates. When we add the subjective probabilities assigned to peace by experts first asked to respond to inevitability curves and the subjective probabilities assigned to war by experts first asked to respond to impossibility curves, the sums will consistently exceed 1.0. Second, there will be a twilight-zone period during which experts who first complete inevitability curves will deem peace inevitable, but experts who first complete impossibility curves will judge war still to be possible. The rationale for the between-group nature of the comparisons is worth stating explicitly because it underscores the critical advantages of experimentation in this context. Given that the experimental groups were constituted by random assignment and hence should not differ systematically in preexisting attitudes, there is no methodological reason to expect systematically different responses to the logically equivalent inevitability- and impossibility-curve questions. Across conditions, the error variance in responses should be normally distributed around the same "true" population mean of respondents' beliefs about the likelihood of peace or war.

**Methods and Measures.** Pilot groups for experiments 1 and 2 were informally drawn from faculty at two large American universities. Respondents for the actual treatment were then randomly selected from the membership lists of divisions 18 and 19 of the APSA, the Society for Military Historians, and the Society for Historians of American Foreign Relations. All respondents were contacted by mail and were promised complete anonymity and detailed feedback on the purposes of the survey. The response rate was 26%.

Experiment 1 randomly assigned the 76 participants to one of three conditions. First, in the control condition, respondents ( $n = 30$ ) were asked (1) when some form of peaceful resolution of the Cuban missile crisis became inevitable and, having identified a point of no return, to estimate the likelihood of a peaceful resolution for each preceding day of the crisis (thereby creating inevitability curves). (2) They were also asked when all alternative (more violent) endings became

impossible and, having identified an "impossibility" date, to estimate the likelihood of those alternative endings on each preceding day (thereby creating impossibility curves).

Second, in the moderate-salience condition, before making retrospective likelihood judgments, respondents ( $n = 23$ ) judged the plausibility of three close-call scenarios. (1) "If Kennedy had heeded his more hawkish advisors in the initial meetings of October 16, there would have been an American air strike against Soviet missile bases in Cuba, and possibly a follow-up invasion of Cuba." (2) "If at least one Soviet ship either did not receive orders to stop before the blockade line (or, for some reason, disobeyed orders), there would have been a naval clash between American and Soviet forces in the Atlantic that would have resulted in military casualties, raising the possibility of tit-for-tat escalation." (3) "If, in the aftermath of the shooting down of a U.S. reconnaissance plane over Cuba on October 20, Kennedy had agreed to implement his standing order to carry out retaliatory air strikes against Soviet SAM (surface to air missile) sites in Cuba that shot down U.S. aircraft, then the U.S. Air Force would have attacked Soviet anti-aircraft installations, which might have set off tit-for-tat escalation." As in the correlational studies, respondents made three judgments of each scenario on nine-point scales: the ease of imagining that antecedent could have occurred; the likelihood of the hypothesized consequence if the antecedent had occurred; and the long-term effect on history if the hypothesized antecedent and consequence did occur.

Third, in the high-salience condition, respondents ( $n = 23$ ) not only considered the three aforementioned situations but also judged a series of nine additional what-if scenarios that reinforced the antecedents in each of the three close calls. For example, counterfactual arguments 1, 2, and 3 reinforced the plausibility of the antecedents in the fourth counterfactual. (1) "If there had not been someone with the intellectual stature and credibility of Secretary of Defense McNamara to make a credible case for caution, then Kennedy would have followed the advice of his more hawkish advisors." (2) "If one of the newspapers to whom Kennedy had confided details of the Soviet placement of missiles in Cuba had leaked the story, there would have been irresistible public pressure on Kennedy to follow the advice of his more hawkish advisors." (3) "If Kennedy had believed that the United States Air Force could knock out all of the Soviet missiles in a single strike (with no need for a follow-up land invasion), he would have followed the advice of his more hawkish advisors." (4) "If Kennedy had followed the advice of his more hawkish advisors in the initial meetings of October 16, there would have been an American air strike against Soviet missile bases in Cuba, and possibly a follow-up invasion of Cuba." The full text and set-up for the presentation of the antecedent-bolstering arguments is available from the authors on request.

**Retrospective Perceptions of Inevitability and Impossibility.** The order of administration of these questions was always counterbalanced. The inevitability-curve exercise instructions were as follows.

Let's define the crisis as having ended when, on October 29, Kennedy communicated to the Soviet leadership his agreement with Khrushchev's radio message of October 28. At that juncture, we could say that some form of peaceful resolution was a certainty—a subjective probability of 1.0. Going backward in time, day by day, from October 29 to October 16, trace on the graph your perceptions of how the likelihood of a peaceful resolution rose or fell during the 14 critical days of the crisis. If you think the U.S. and U.S.S.R. never came close to a military clash between October 16 and 29, then express that view by assigning consistently high probabilities to a peaceful resolution across all dates (indeed, as high as certainty, 1.0, if you wish). If you think the superpowers were very close to a military conflict throughout the crisis, then assign consistently low probabilities to a peaceful resolution across all dates. Finally, if you think the likelihood of a peaceful resolution waxed and waned day to day, then assign probabilities that rise or fall in accord with your intuitions about how close the U.S. and U.S.S.R. came to a military clash at various junctures. To start, we have set the subjective probability of peace at 1.0 (certainty) for October 29, marking the end of the crisis.

The impossibility-curve instructions were similar, except that the starting point was the subjective probability of 0.0 assigned to October 29 to signify that alternative, more violent outcomes had become impossible. Experts were then asked to go backward in time, day by day, from October 29 to October 16, and trace on the graph their perceptions of how the likelihood of those more violent outcomes waxed and waned.

**Findings.** The initial analyses involved a  $3 \times 2 \times 13$  fixed-effects, unweighted-means analysis of variance that crossed three levels of the between-subjects experimental manipulation (control, moderate, and high salience), two levels of the individual-difference classification variable (low versus high need for closure), and thirteen levels of the repeated-measures factor that corresponded to the days of the crisis. Contrary to expectation, the moderate and high conditions did not differ on either inevitability or impossibility curves (both  $F$ 's  $< 1$ ). We attribute this null result to a methodological shortcoming: Respondents reported being rather overwhelmed by the number of judgments required in the high-salience condition, and fatigue may have attenuated any further effect that exposure to additional counterfactual scenarios might have had.

To simplify analysis, therefore, we collapsed the moderate and high groups into a single salient condition. Follow-up analyses, now taking the form of a  $2 \times 2 \times 13$  analysis of variance, revealed the predicted second-order interaction: Inevitability curves rose more slowly over time among those with lower need for closure assigned to the salient condition,  $F(12, 908) = 6.74, p < .01$ . The predicted mirror-image second-order interaction emerged on the impossibility curves  $F(12, 908) = 5.33, p < .01$ , which is not surprising, given that the measures were highly corre-

lated,  $r = 0.76$ . Figures 1 and 2 clearly show that the distinctive functional forms of the inevitability and impossibility curves of low-need-closure respondents in the salient condition drive both interactions.

As expected, within-subjects comparisons reveal that when experts completed an inevitability curve and immediately thereafter an impossibility curve—that is, when binary complementarity was transparently at stake—subjective probabilities of peace and war summed to approximately 1.0 ( $\bar{x} = 1.04$ ). Systematic violations of binary complementarity emerged, however, when we made more subtle between-group comparisons. For instance, when we add the subjective probability of peace assigned by experts who first completed inevitability curves to the subjective probability of war assigned by experts who first completed impossibility curves, the average sum across dates is 1.19. This value is significantly different from what we obtain by adding the probability of war and peace judgments of the two groups of experts who completed their inevitability or impossibility curves in the second position: The average sum across dates = 0.90,  $F(1, 71) = 10.32, p < .01$ . There was, however, no evidence for the twilight-zone-period hypothesis that the experts who responded first to either inevitability or impossibility curves could be “lured” into assigning probability values that implied the existence of a period during which peace was inevitable (1.0) but war had not yet become impossible (0.0),  $\bar{x}$  impossibility date of war = October 27.5 and  $\bar{x}$  inevitability date of peace = 26.9,  $F(1, 71) = 2.68, p < .15$ .

## Experiment 2: Unpacking Alternative Outcomes of the Cuban Missile Crisis

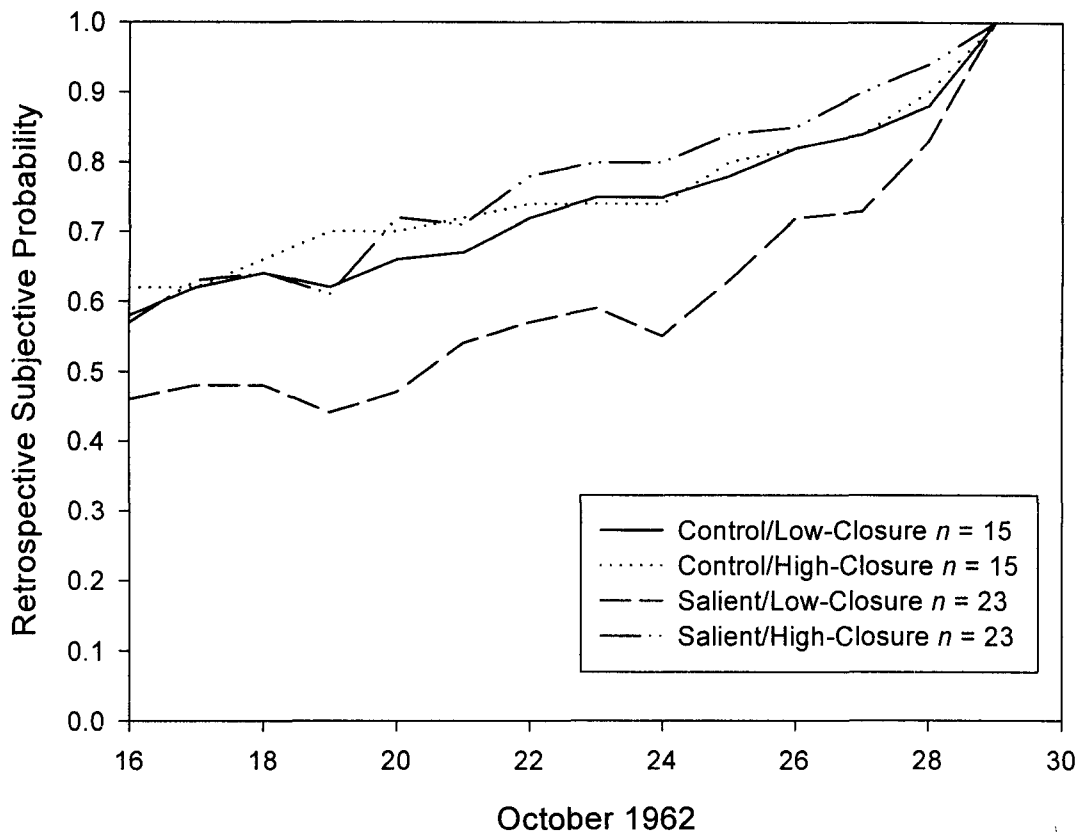
Skeptics can argue that in Experiment 1 respondents were confronted with an elaborate battery of mutually reinforcing counterfactuals that made alternative histories unfairly vivid and left little room for deterministic rejoinders. It also can be argued that norms of politeness made experts reluctant to dismiss all the researchers' what-if scenarios as errant nonsense. Experiment 2 eliminates both objections by shifting the spotlight to the power of entirely self-generated counterfactual scenarios to alter perceptions of historical contingency.

**Guiding Theory.** Consider again forecasts of which league, division, or team will win a sports championship. Tversky and Fox (1995) demonstrate that the subjective probabilities people assign to binary complements at the league level (East vs. West) generally sum to 1.0, but the subjective probabilities assigned to progressively more detailed or unpacked outcomes—the prospects of divisions within leagues and teams within divisions—typically exceed 1.0 and occasionally even 2.0. Forecasters find it easier to generate evidential support for a particular team winning than for several different teams winning.

In support theory, it is the ease with which these reasons come to mind, their availability, that determines the subjective feeling of support for, and subject-



FIGURE 1. Inevitability Curves from Experiment 1



Note: The figure displays inevitability curves from experts with low and high need for closure in the control and salient conditions of Experiment 1. The rate of rise toward 1.0 indicates the degree to which experts perceived the likelihood of some form of peaceful resolution of the Cuban missile crisis as increasingly likely with the passage of time, with the value of 1.0 signifying inevitability.

tive probability of, outcomes. The result can be massive “subadditivity.” The cumulative probabilities assigned to the exhaustive and exclusive components of the whole set exceed 1.0, which violates the extensionality axiom of probability theory. If people were to back up their unpacked bets with actual money, they would be quickly transformed into money pumps. It is, after all, logically impossible for each of four teams within an eight-team division to have a 0.4 chance of winning the championship the same year.

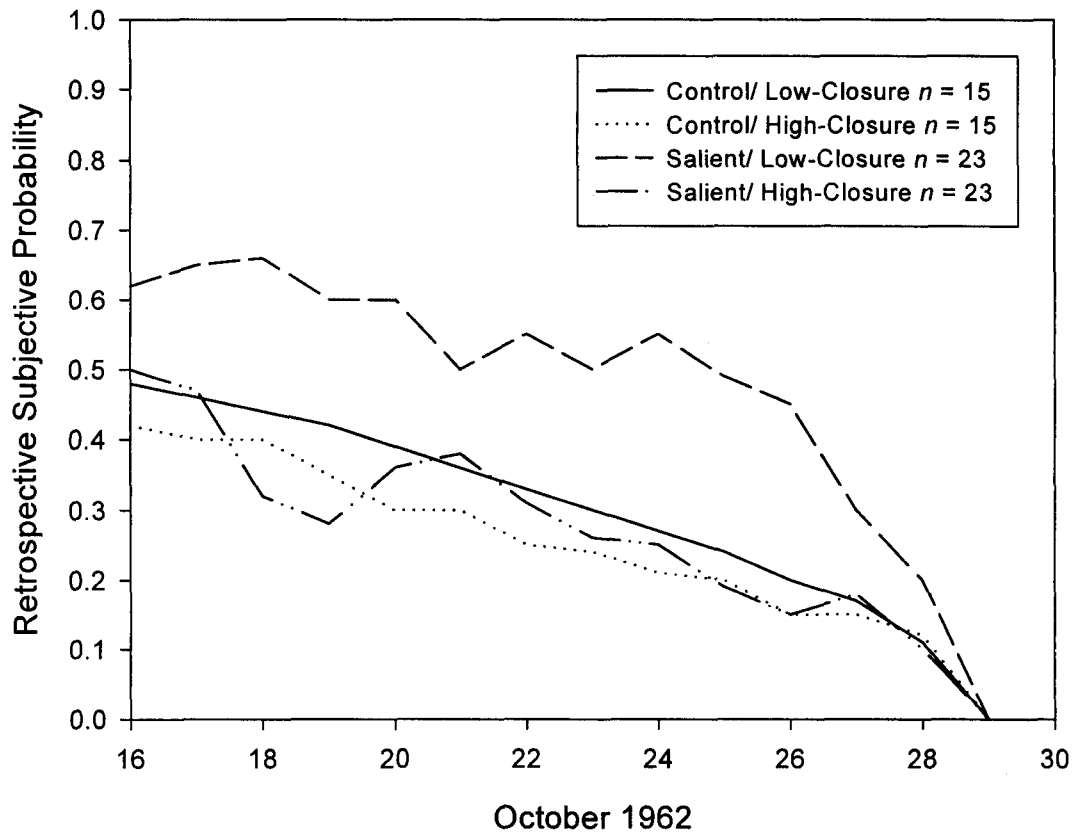
Unpacking manipulations are understandably viewed as sources of cognitive bias in subjective probability judgments of possible futures. They stimulate people to find too much support for too many possibilities. Yet, such manipulations may help reduce bias in subjective probability judgments of possible pasts via exactly the same mechanism. The key difference is that judgments of possible pasts, unlike those of possible futures, are already contaminated by the powerful certainty of hindsight. Experimental work shows that as soon as people learn which of a number of once-deemed possible outcomes happened, they quickly assimilate that knowledge into their cognitive structure and have a hard time recapturing their ex ante state of uncertainty (Hawkins and Hastie 1990). Mental exercises that involve unpacking sets of possible pasts

should have the net effect of checking the hindsight bias by bringing back to psychological life counterfactual possibilities that people long ago buried with deterministic “I-knew-it-had-to-be” thinking.

**Hypotheses.** Drawing on support theory, we hypothesize that experts who are encouraged to unpack the set of more violent endings of the Cuban missile crisis into progressively more differentiated subsets will find support for those alternative outcomes. As a result, their inevitability curves will rise more slowly and their impossibility curves will fall less rapidly than those of experts who judge the entire set of possibilities as a whole. It is also expected that experts in the unpacking condition, especially those with low need for closure, will display stronger subadditivity effects (cumulative subjective probabilities exceeding 1.0) than the holistic group.

**Research Design, Method, and Logic of Analysis.** The 64 respondents in Experiment 2 were drawn from the same subject population as Experiment 1 and recruited in the same mail survey. Respondents were randomly assigned to one of two groups. The control group ( $n = 30$ ) simply responded to the perceptions-of-inevitability and perceptions-of-impossibility items, as in Experiment 1. The other group ( $n = 34$ ) was asked to

FIGURE 2. Impossibility Curves from Experiment 1



Note: The figure displays impossibility curves from experts with low and high need for closure in the control and salient conditions of Experiment 1. The rate of decline toward zero indicates the degree to which experts perceived the likelihood of alternative, more violent endings of the Cuban missile crisis as decreasingly likely with the passage of time, with zero signifying impossibility.

consider (1) how the set of more violent endings of the Cuban missile crisis could be disaggregated into subsets in which violence remained localized or spread outside the Caribbean, (2) in turn differentiated into subsets in which violence claimed fewer or more than 100 casualties, and (3) for the higher casualty scenario, still more differentiated into a conflict either limited to conventional weaponry or extending to nuclear. Respondents generated impossibility curves for each of the six specific subsets of more violent scenarios as well as a single inevitability curve for the overall set of peaceful outcomes.

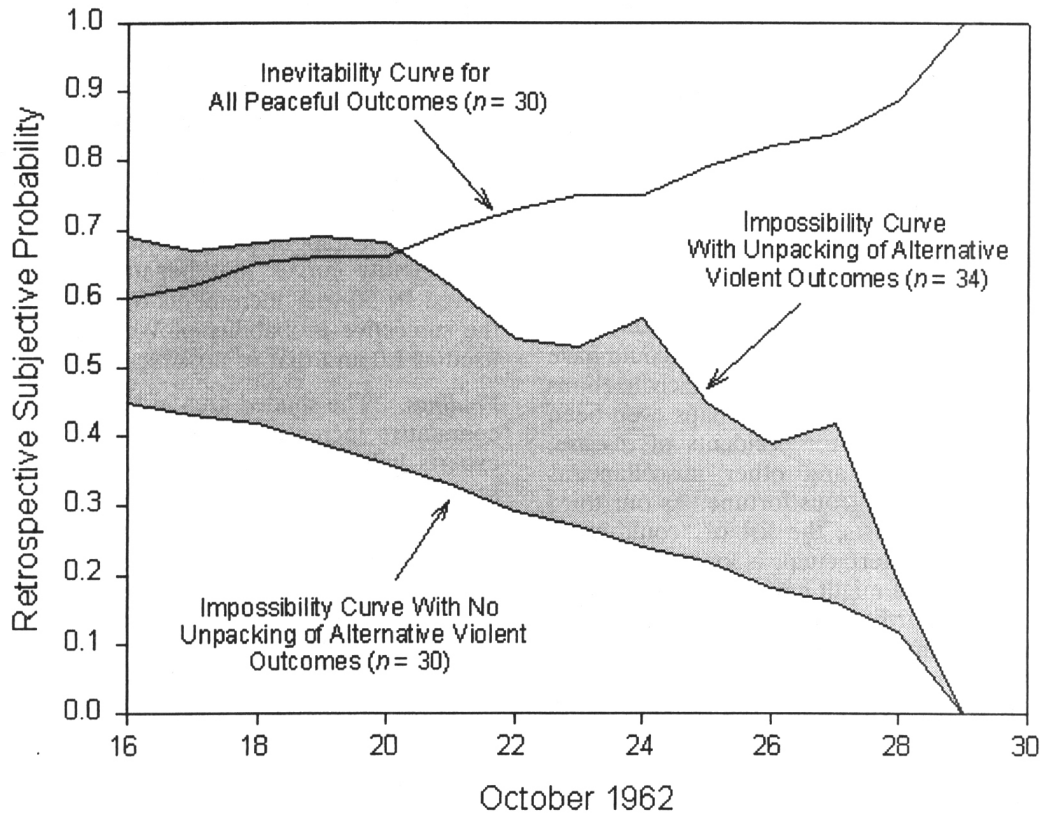
**Findings.** The results again reveal that how we pose historical questions shapes how we answer them. Figure 3 illustrates the power of unpacking questions. The shaded area represents the cumulative increase in the subjective probability that experts believe counterfactual alternatives once possessed, an increase that was produced by asking experts to generate impossibility curves not for the abstract set of more violent outcomes (lower curve) but for each of the six specific subsets of those outcomes (upper curve). The analysis of variance took the form of a fixed-effects, unweighted means 2 (control versus unpacking)  $\times$  2 (low versus high need for closure)  $\times$  13 (days of crisis) design.

Consider the impossibility-curve dependent variable.

(Inevitability-curve results were again highly correlated,  $r = .71$ , and largely redundant for these hypothesis-testing purposes.) Analysis revealed the predicted main effects for unpacking ( $F(1, 58) = 7.89, p < .05$ ) and need for closure ( $F(1, 58) = 5.05, p < .05$ ), as well as the expected tendency for the impossibility curve of respondents with low need for closure to fall more slowly than that of high-need respondents in the unpacking condition ( $F(1, 58) = 4.35, p < .05$ ). In addition, two unexpected tendencies emerged: Unpacking effects diminished toward the end of the crisis ( $F(12, 718) = 7.31, p < .05$ ), as did differences between low- and high-closure respondents ( $F(12, 718) = 5.02, p < .05$ ). Experts, even low-need-closure experts unpacking possibilities, saw less and less wiggle room for rewriting history as the end approached.

There was also support for the hypothesis that low-closure experts in the unpacking condition will exhibit the strongest subadditivity effects (probability judgments of exhaustive and exclusive sets of possibilities summing to more than 1.0). Averaged across dates, their combined inevitability and impossibility judgments summed to 1.38, which was significantly greater than the sum for low-closure experts in the control group ( $\bar{x} = 1.12$ ) or for high-closure experts in

**FIGURE 3. Inevitability and Impossibility Curves from Experiment 2**



*Note:* The figure presents inevitability and impossibility curves for the Cuban missile crisis. The inevitability curve displays gradually rising likelihood judgments of some form of peaceful resolution. The lower impossibility curve displays gradually declining likelihood judgments of all possible more violent endings. The higher impossibility curve was derived by adding the experts' likelihood judgments of six specific subsets of more violent possible endings. Adding values of the lower impossibility curve to the corresponding values of the inevitability curve yields sums only slightly above 1.0. Inserting values from the higher impossibility curve yields sums well above 1.0. The shaded area represents the cumulative effect of unpacking on the retrospective subjective probability of counterfactual alternatives to reality.

either the unpacking condition ( $\bar{x} = 1.18$ ) or control group ( $\bar{x} = 1.04$ ) ( $F(1, 58) = 9.89, p < .05$ ). Again, there was little support for the twilight-zone-period hypothesis. The longest time during which experts judged peace inevitable ( $\bar{x}$  inevitability date = Oct. 27.2) but war not yet impossible ( $\bar{x}$  impossibility date = Oct. 28.1) emerged in judgments within the unpacking condition, and even this difference fell short statistically ( $F(1, 58) = 3.03, p < .10$ ).

The curve-fitting results also underscore the power of counterfactual thought experiments to transform our understanding of the past. Simple linear equations capture large proportions of the variance in retrospective-likelihood judgments of the undifferentiated sets of peaceful outcomes (82%) and more violent alternatives (84%). The past appears to be a smooth linear progression toward the observed outcome. By contrast, the past looks more like a random walk, albeit around a discernible trend, from the perspective of low-closure experts who unpacked the set of more violent outcomes. A convoluted fourth-order polynomial equation is necessary to explain the same proportion of variance in their retrospective likelihood judgments, a function that rises and falls at three junctures.

The power of unpacking is also revealed by cross-condition comparisons of correlations between theoretical beliefs, such as the robustness of nuclear deterrence, and reactions to close-call counterfactuals that move the missile crisis toward war. The correlation is greater in the control condition than in the unpacking condition ( $r(28 \text{ df}) = 0.61$  versus  $r(32 \text{ df}) = 0.27$ ). This drop is consistent with the notion that, under unpacking, observers shift from a theory-driven, covering-law mode of thinking to a more idiographic, case-by-case mode.

### Experiment 3: Unmaking the West

**Guiding Theory.** Scholars have long pondered how a small number of Europeans, working from the superficially unpromising starting point of 1000 A.D. or 1200 A.D. or even 1400 A.D., managed in relatively a few centuries to surpass all other peoples on the planet in wealth and power. Not surprisingly, there is a wide range of opinion. At one pole are determinists, who view history as an efficient process of winnowing out maladaptive forms of social organization and who believe that the triumph of capitalism has long been in



the cards. The key advantages of European polities allegedly included more deeply rooted legal traditions of private property and individual rights, a religion that encouraged worldly achievement, and a fractious multistate system that prevented any single power from dominating all others and halting innovation at the reactionary whim of its ruling elite (McNeill 1982).

At the other pole are the antideterminists. To adapt Gould's (1995) famous thought experiment, they believe that if we could rerun world history thousands of times from the starting conditions that prevailed as recently as 1400 A.D., European dominance would be one of the least likely outcomes. These scholars decry "Eurocentric triumphalism" and depict the European achievement as a precarious one that easily could have unraveled at countless junctures. Other civilizations could have checked the West and perhaps even been contenders themselves but for accidents of disease, weather, bad leadership, and other miscellaneous slings and arrows of outrageous fortune. As our third correlational study suggests, the list of "could-have-been-a-contender" counterfactuals is long. South Asia and perhaps East Africa might have been colonized by an invincible Chinese armada in the fifteenth century if only there had been more support in the imperial court for technological innovation and territorial expansion. Europe might have been Islamicized in the eighth century if the Moors had cared to launch a serious invasion of France. If not for Genghis Khan dying in a nick of time, European civilization might have been devastated by Mongol armies in the thirteenth century.

Within the antideterministic framework, thought experiments become exercises in ontological egalitarianism, an effort to restore dignity to those whom history has eclipsed by elevating possible worlds to the same moral and metaphysical status as the actual world (Tetlock n.d.). Thought experiments are the only way left to even the score, an observation ironically reminiscent of the Marxist historian E. H. Carr's (1961) dismissal of anti-Bolsheviks as sore-losers who, from dreary exile, contemplated counterfactuals that undid the Russian Revolution. But now the gloaters, claiming historical vindication for their ideological principles, are on the Right, and the brooders, absorbed in wistful regret, are on the Left.

**Hypotheses.** The hypotheses parallel those for Experiment 2, except now the focal issue is not the Cuban missile crisis but the rise of Western civilization to global hegemony (a massively complex historical transformation that stretches over centuries, not days). Once again, unpacking is expected to inflate the perceived likelihood of counterfactual possibilities and to produce subadditivity effects, especially for respondents with low need for closure.

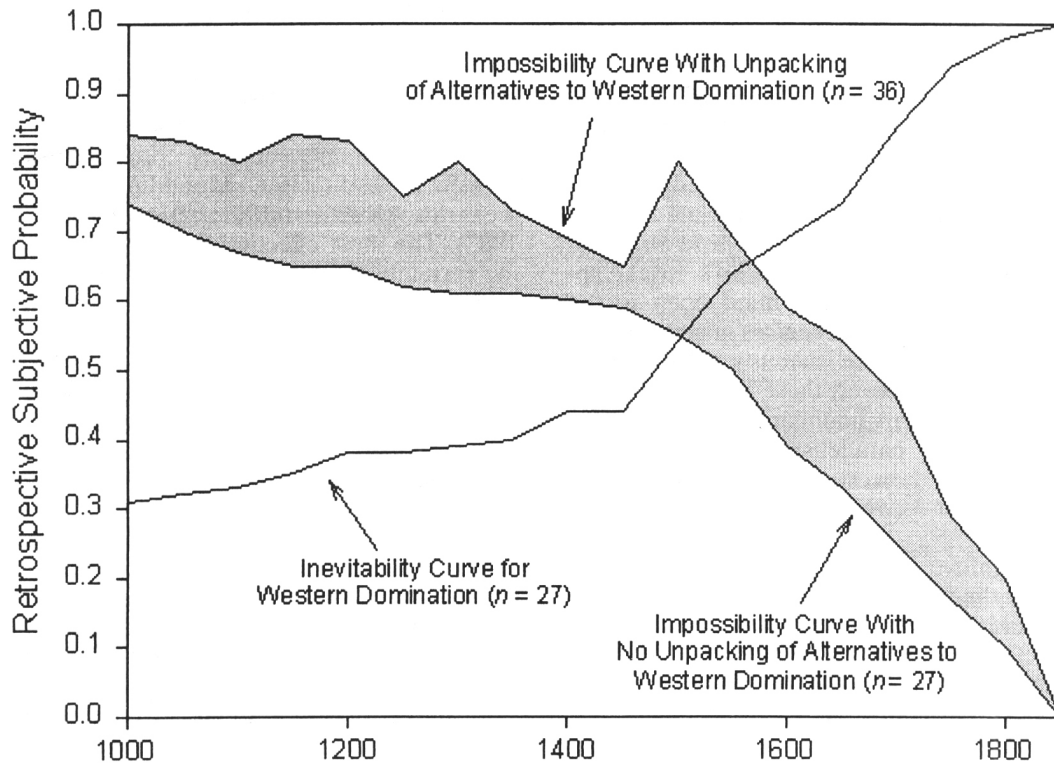
**Research Design, Methods, and Measures.** Experiment 3 draws on the same respondents and uses the same mail survey as the third correlational study. The experiment has only two conditions. The no-unpacking control group ( $n = 27$ ) generated inevitability curves for some form of Western geopolitical domination and impossibility curves for the set of all possible alterna-

tives to that domination (order counterbalanced). The intensive unpacking group ( $n = 36$ ) was first asked to unpack the set of all possible alternatives to Western domination into progressively more detailed subsets. These began with classes of possible worlds in which no region achieved global hegemony (either because of a weaker Europe or stiffer resistance from outside Europe) and moved on to classes of possible worlds in which a non-Western civilization achieved global hegemony (China, Islam, the Mongols, or a less familiar alternative). Experts then completed inevitability and impossibility curves that began with 1000 A.D. and moved by 50-year increments to 1850 A.D. (for which the subjective probability of Western dominance was fixed at 1.0 and that of possible alternatives at 0.0).

**Findings.** The shaded area in Figure 4 represents the cumulative increase in the subjective probability that experts believe counterfactual alternatives once possessed, an increase that was produced by asking experts to generate impossibility curves not for the abstract set of alternatives to Western domination (lower curve) but for each of the six specific subsets of alternatives (upper curve). The analysis of variance took the form of a fixed-effects, unweighted means 2 (no unpacking versus unpacking conditions)  $\times$  2 (low versus high closure)  $\times$  17 (50-year increments between 1000 and 1850) mixed-factorial design. It reveals a significant unpacking effect ( $F(1, 58) = 6.77, p < .05$ ) and a significant interaction between unpacking and cognitive style: The differences between low- and high-closure respondents grew more pronounced when they were asked to perform the unpacking exercise ( $F(1, 58) = 4.88, p < .05$ ). The same two unexpected but readily interpretable tendencies emerged as in Experiment 2. That is, as dates approached the end of the time series, now 1800, experts who unpacked alternatives to Western dominance saw less potential for rewriting history ( $F(16, 1018) = 4.88, p < .01$ ), and the gap between low- and high-closure respondents narrowed ( $F(16, 1018) = 4.02, p < .05$ ). Again, observers saw less likelihood for counterfactually altering outcomes toward the end of the specified historical process.

Also as in Experiment 2, low-closure experts in the unpacking condition were most likely to assign subjective probabilities that were subadditive, violated extensionality, and summed to well above 1.0. Averaged across dates, their judgments about inevitability and impossibility summed to 1.41, which was significantly greater than the sum for low-closure experts in the control group ( $\bar{x} = 1.09$ ) or for high-closure experts in the control group ( $\bar{x} = 1.03$ ) or unpacking condition ( $\bar{x} = 1.21$ ) ( $F(1, 58) = 4.67, p < .05$ ). A twilight-zone period also emerged; in contrast to the temporally compacted missile crisis, there is now a sufficient range in responses to permit significant effects. Focusing just on the control condition, we compared the average date for the inevitability of Western dominance to the average date for the impossibility of alternatives to Western dominance. The result was a twilight-zone period bounded by the inevitability date

**FIGURE 4. Inevitability and Impossibility Curves from Experiment 3**



Note: The figure presents inevitability and impossibility curves for the rise of the West. The inevitability curve displays gradually rising likelihood judgments of some form of Western geopolitical dominance. The lower impossibility curve displays gradually declining likelihood judgments of all possible alternatives to Western dominance. The upper impossibility curve was derived by adding experts' likelihood judgments of six subsets of alternatives to Western dominance. Adding values of the lower impossibility curve to corresponding inevitability-curve values yields sums only slightly above 1.0. Inserting values from the upper impossibility curve yields sums well above 1.0. The shaded area represents the cumulative effect of unpacking on the retrospective subjective probability of counterfactual alternatives to reality.

of 1731 and the impossibility date of 1749 ( $F(1, 28) = 4.21, p < .05$ ). This period for the unpacking condition was bounded, respectively, by 1751 and 1787, which not only is a significant difference in itself ( $F(1, 32) = 8.43, p < .01$ ) but also is significantly longer ( $F(1, 58) = 4.36, p < .05$ ).

In terms of curve fitting, a fifth-order polynomial equation was necessary to capture 80% of the variance in the numerous ups and downs in average-perceptions of the likelihood of unpacked outcomes, whereas a simple linear equation sufficed for the undifferentiated set. Unpacking “messes up” our understanding of the past, transforming what had once been a smooth progression toward a preordained outcome into a far more erratic and unpredictable journey. Unpacking also undermines the power of abstract covering laws to constrain our perceptions of specific historical possibilities; correlations fell from 0.63 in the control condition to 0.25 in the unpacking conditions.

**GENERAL DISCUSSION**

Our results do not tell us who is right about this or that historical controversy. Proponents of the covering-law approach can argue that high correlations between abstract theoretical beliefs and specific opinions about what is possible at given times and places are fully

justified and that low-preference-for-parsimony experts who fail to make tight conceptual connections between the two levels are just confused and sloppy thinkers. As for the results of our experiments, these proponents can argue that unpacking manipulations simply lead historical observers into the inferential equivalent of wild goose chases that encourage them to assign far too much subjective probability to far too many scenarios. The end result is logical absurdities, such as “*x* is inevitable,” but “alternatives to *x* remain possible.” There is nothing admirably open-minded about incoherence.

Critics of the covering-law approach can respond that the high correlations between abstract theoretical beliefs and specific historical ones are a warning sign of an excessively theory-driven style of thinking about history. Insofar as unpacking manipulations prove an effective method of opening closed minds by reminding theory-driven thinkers of how riddled with indeterminacy history is, then so much the better, and do not fret that unpacking induces internal inconsistencies in belief systems. Unpacking reveals previously hidden cognitive contradictions and sets a constructive agenda for rethinking. A foolish consistency, these critics might admonish us, is the hobgoblin of little minds.

A cognitive approach to historical reasoning can never resolve the question of whether, in any given

historical case, we have struck the right balance between theory-driven and imagination-driven modes of information processing. A cognitive approach can, however, enhance the quality of intellectual debate in four tangible ways.

First, it identifies systematic individual differences in the relative importance that observers place on achieving an integrative view of the past. Respondents with a preference for closure (who resemble the hedgehogs in Berlin's famous taxonomy) tried to assimilate the past into their favored deductive system. Those with less need for closure (who resemble Berlin's foxes) approached the past in a frame of mind open to the possibility that a potpourri of processes may have been involved, and an equally miscellaneous set of outcomes could have resulted. Moreover, these individual differences cut across disciplinary boundaries, a point underscored by the striking parallelism in the results of Experiment 2 (which drew mostly on political scientists) and Experiment 3 (which drew mostly on historians).

Second, a cognitive approach deepens our understanding of how the framing of historical questions can shape where and how we look for answers. Logically equivalent factual and counterfactual framings of the same historical issue elicit contradictory probabilistic intuitions about what had to be and what might have been. Our experiments also demonstrate the importance of the specificity of the question: The more detailed the unpacking of counterfactual alternatives, the greater is the tendency for the whole class of alternatives to be judged less likely than the sum of its exclusive and exhaustive components (subadditivity). The challenge that these findings pose to rationality should not be understated. Unpacking effects violate a core assumption not only of Bayesian but also of all formal logical models of belief, namely, the extensionality principle, which asserts that classes of events with the same content should be assigned the same probability (Tversky and Fox 1995).

Third, a cognitive approach suggests methods of pitting cognitive biases against each other, the mental equivalent of fighting fire with fire, by checking the excesses of convergent, theory-driven thinking with divergent imagination-driven thinking, and vice versa. The preeminent theory-driven bias in historical reasoning is, as many laboratory studies now attest, certainty of hindsight (Hawkins and Hastie 1990). Once people learn the outcome of an historical process—how the stock market finished the year or whether a crisis was resolved peacefully—they have difficulty recalling how uncertain they once were and exaggerate the degree to which “they knew it all along.” The most influential cognitive explanation of the hindsight effect attributes it to the “automaticity” of theory-driven thought, to the rapidity with which people assimilate known outcomes into their favorite covering laws. In the process, they demote once possible, even probable, futures to the status of implausible historical counterfactuals.

Knowledge of past possibilities need not, however, be lost forever. The “debiasing” literature points to experimental interventions that can help people recon-

struct latent knowledge of historical junctures at which they once thought events could have taken a different turn (Tetlock and Belkin 1996). The potential correctives draw on two well-replicated effects widely considered judgmental biases themselves: the tendency to inflate the perceived likelihood of vivid, easily imagined, scenarios, coupled with the tendency to underestimate how rapidly subjective probabilities should diminish as a function of adding contingencies to scenarios (Koehler 1991; Tversky and Kahneman 1983). The most effective manipulations for attenuating certainty of hindsight make use of these effects by asking people to imagine, in florid detail, alternative paths that history could have taken.

Fourth, a cognitive approach warns us not to go too far in the quest to open closed minds. Laboratory work on mental simulation shows that people can get carried away when they start imagining alternative “possible worlds” (Koehler 1991). To check runaway unpacking effects, people need plausibility pruners to cut off speculation that otherwise would grow like topsy beyond the bounds of probability. These elimination criteria inevitably reflect covering-law beliefs about what counts as a plausible cause-effect linkage. Indeed, the cognitive approach also warns us not to mystify or romanticize imagination and treat it as somehow superior to theory-driven thinking. Alice-in-Wonderland scenarios aside, most imaginative thinking imports theory-driven schemes for understanding cause and effect in hypothetical worlds. The distinction between the two modes is a matter of degree: theory-driven cognition is convergent, relying on a single deductive framework to explain what actually happened, whereas imaginative cognition is divergent, drawing on a variety of frameworks to explore not only what happened but also what could have happened.

If this account is correct, historical observers confront a perplexing trade-off between theory-driven and imagination-driven modes of making sense of the past. Theory-driven strategies confer the benefits of explanatory closure and parsimony but desensitize us to nuance, complexity, contingency, and the possibility our theory is wrong. Imagination-driven strategies sensitize us to possible worlds that could have been, but the price is increased confusion and even incoherence. The challenge is to strike a reasonable balance, a reflective equilibrium, between the conflicting intuitions primed by factual framings of historical questions that activate narrowly theory-driven thinking and those primed by counterfactual framings that engage our imaginations and activate a wider range of theories.

Absent compelling grounds for preferring one or the other mode of information processing, the prudent stance is to define good judgment as an iterative process of checking convergent theory-driven thinking, which focuses on explaining why what was had to be, with divergent imagination-driven thinking, which focuses on what could have been, and then putting the whole procedure into reverse so that theory-driven thinking can be deployed to prevent us from spinning out of imaginative control. There will, of course, be no single, well-defined equilibrium solution. But it is rea-



sonable to expect that the process—as it unfolds not just in the mind of one observer but in ostensibly self-correcting scholarly communities—will yield plausibility ranges of equilibrium solutions. These will be anchored at one end by those with the most hedgehog-like confidence in covering laws and at the other by those with the most fox-like curiosity about the paths events could have taken. This range of reflective-equilibrium solutions should be preferred to lopsided solutions that give priority to either theory-driven or imaginative modes of thinking.

## CONCLUSION

Objectivity was the epistemological rock on which professional societies of historians and social scientists were founded in the late nineteenth century. The disciplinary mandate was to move closer, by successive approximations, toward the truth, a truth unadorned by apologetic or mocking quotation marks (Novick 1991). Well before century's end, however, objectivity was under siege as doubts grew about positivism and about the feasibility of drawing sharp distinctions between observer and observed, between facts and values, and even between facts and fiction. Constructivist and reflexivist epistemologies, which depicted truth as perspectival, not unitary, gained in influence.

Our research provides conceptual ammunition to both constructivist opponents and neopositivist proponents of objectivity. Constructivists can take heart from our findings that theoretical preconceptions are potent moderators of what “observers” see as possible or impossible in event sequences. They can also point to repeated demonstrations that how we pose historical questions, even logically equivalent factual and counterfactual framings of the same question, shapes the conclusions we draw about what had to be or could have been. But positivists need not surrender. The discipline that provided the concepts and methods underlying the studies reported here—cognitive psychology—is itself a neopositivist enterprise. There are many precedents in the history of science in which the scientific method has highlighted potential biases and correctives in scientific practice. Defenders of “that noble dream,” objectivity, can quite defensibly argue that identifying judgmental biases is a prerequisite to correcting them through disciplined exercises in reflective equilibrium.

## REFERENCES

- Berlin, Isaiah. 1997. “The Hedgehog and the Fox.” In *The Proper Study of Mankind*, ed. Isaiah Berlin. New York: Farrar, Straus and Giroux. Pp. 436–98.
- Carr, E. H. 1961. *What Is History?* London: Penguin Books.
- Dray, William H. 1989. *On History and Philosophers of History*. Leiden, NY: Brill.
- Fearon, James. 1991. “Counterfactuals and Hypothesis Testing in Political Science.” *World Politics* 43 (April): 474–84.
- Fogel, Robert. 1964. *Railroads and American Economic Growth: Essays in Econometric History*. Baltimore, MD: Johns Hopkins University Press.
- Goldstone, Jack. N.d.. “Europe’s Peculiar Path: The Unlikely Transition to Modernity.” *Unmaking the West: Alternative Histories of Counterfactual Worlds*, ed. Philip E. Tetlock, Richard Ned Lebow, and Geoffrey Parker. New York: Columbia University Press. Forthcoming.
- Gould, Stephen Jay. 1995. *Dinosaur in a Haystack: Reflections in Natural History*. New York: Harmony Books.
- Hawkins, Scott, and Reid Hastie. 1990. “Hindsight: Biased Judgment of Past Events after the Outcomes Are Known.” *Psychological Bulletin* 107 (21): 311–27.
- Hempel, Carl G. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.
- Jervis, Robert. 1976. *Perception and Misperception in International Politics*. Princeton, NJ: Princeton University Press.
- Kahneman, Daniel, and Dale Miller. 1986. “Norm Theory: Comparing Reality to Its Alternatives.” *Psychological Review* 93 (April): 136–53.
- Koehler, Derek. 1991. “Explanation, Imagination, and Confidence in Judgment.” *Psychological Bulletin* 110 (3): 499–519.
- Kruglanski, Arie W., and Donna M. Webster. 1996. “Motivated Closing of the Mind: ‘Seizing’ and ‘Freezing.’” *Psychological Review* 103 (2): 263–8.
- Landes, David S. 1998. *The Wealth and Poverty of Nations*. New York: W. W. Norton.
- Lebow, Richard Ned. 2001. “Contingency, Catalysts and International System Change.” *Political Science Quarterly* 115 (Winter): 591–616.
- Lewis, David K. 1973. *Counterfactuals*. Cambridge, MA: Harvard University Press.
- Lustick, Ian S. 1996. “History, Historiography, and Political Science: Multiple Historical Records and the Problem of Selection Bias.” *American Political Science Review* 90 (September): 605–18.
- McCloskey, Donald N. 1991. “History, Differential Equations, and the Problem of Narration.” *History & Theory* 36 (3): 37–62.
- McNeill, William H. 1982. *Pursuit of Power: Technology, Armed Force, and Society since A.D. 1000*. Chicago: University of Chicago Press.
- Mokyr, Joel. N.d. “King Kong and Cold Fusion.” In *Unmaking the West: Counterfactual Explorations of Alternative Worlds*, ed. Philip E. Tetlock, Richard Ned Lebow, and Geoffrey Parker. New York: Columbia University Press. Forthcoming.
- Neustadt, Richard E., and Ernest R. May. 1986. *Thinking in Time: The Uses of History for Decision-Makers*. New York: Free Press.
- Novick, Peter. 1988. *That Noble Dream: The Objectivity Question and the American Historical Profession*. New York: Cambridge University Press.
- Sagan, Scott, and Kenneth Waltz. 1995. *The Spread of Nuclear Weapons: A Debate*. New York: W. W. Norton.
- Simon, Herbert A. 1996. *The Sciences of the Artificial*. Cambridge, MA: MIT Press.
- Suedfeld, Peter, and Philip E. Tetlock. 2001. “Cognitive Styles.” In *Blackwell International Handbook of Social Psychology: Intra-Individual Processes*, vol 1, ed. Abraham Tesser and Norbert Schwartz. London: Blackwell. Pp. 282–304.
- Tetlock, Philip E. 1998. “Close-Call Counterfactuals and Belief System Defenses: I Was Not Almost Wrong but I Was Almost Right.” *Journal of Personality and Social Psychology* 75 (2): 230–42.
- Tetlock, Philip E. 1999. “Theory-Driven Reasoning about Possible Pasts and Probable Futures: Are We Prisoners of Our Preconceptions?” *American Journal of Political Science* 43 (April): 335–66.
- Tetlock, Philip E. N.d. “The Logic and Psycho-Logic of Counterfactual Thought Experiments in the Rise-of-the-West Debate.” In *Unmaking the West: Counterfactual Explorations of Alternative Worlds*, ed. Philip E. Tetlock, Richard Ned Lebow, and Geoffrey Parker. New York: Columbia University Press. Forthcoming.
- Tetlock, Philip E., and Aaron Belkin. 1996. *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives*. Princeton, NJ: Princeton University Press.
- Tversky, Amos, and Craig Fox. 1995. “Weighting Risk and Uncertainty.” *Psychological Review* 102 (2): 269–83.
- Tversky, Amos, and Daniel Kahneman. 1983. “Extensional versus Intuitive Reason: The Conjunction Fallacy as Probability Judgment.” *Psychological Review* 90 (2): 292–315.
- Vasquez, John A. 1997. “The Realist Paradigm and Degenerative versus Progressive Research Programs: An Appraisal of Neotraditional Research on Waltz’s Balancing Proposition.” *American Political Science Review* 91 (December): 899–913.