

This article was downloaded by: [Lustick, Ian S.]

On: 2 March 2011

Access details: Access Details: [subscription number 934130590]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Critical Review

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t778142998>

## TETLOCK AND COUNTERFACTUALS: SAVING METHODOLOGICAL AMBITION FROM EMPIRICAL FINDINGS

Ian S. Lustick<sup>a</sup>

<sup>a</sup> Department of Political Science, University of Pennsylvania, Philadelphia, PA

Online publication date: 26 February 2011

**To cite this Article** Lustick, Ian S.(2010) 'TETLOCK AND COUNTERFACTUALS: SAVING METHODOLOGICAL AMBITION FROM EMPIRICAL FINDINGS', *Critical Review*, 22: 4, 427 – 447

**To link to this Article:** DOI: 10.1080/08913811.2010.541698

**URL:** <http://dx.doi.org/10.1080/08913811.2010.541698>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

TETLOCK AND COUNTERFACTUALS:  
SAVING METHODOLOGICAL AMBITION  
FROM EMPIRICAL FINDINGS

ABSTRACT: *In five works spanning a decade, Philip E. Tetlock's interest in counterfactuals has changed. He began with an optimistic desire to make social science more rigorous by identifying best practices in the absence of non-imagined controls for experimentation. Soon, however, he adopted a more pessimistic analysis of the cognitive and psychological barriers facing experts. This shift was brought on by an awareness that experts are not rational Bayesians who continually update their theories to keep up with new information; but instead are affected by political, cognitive, and psychological heuristics, including hindsight bias, cognitive conservatism, and the fundamental attribution error. But techniques of computational simulation—involving the rigorous production of large numbers of counterfactual worlds—make it possible to mitigate both problems that Tetlock identifies: that history, produced only once, is a lousy teacher; and that humans, with their collection of non-Bayesian heuristics, are lousy pupils. Tetlock was wrong to reject this approach as theoretically promising but rhetorically and practically impractical.*

In the film *Sliding Doors*, the audience is shown two different lives. If Helen (Gwyneth Paltrow) makes her train and squeezes through the sliding doors in time, she will discover her boyfriend in bed with another woman—but meet a wonderful guy on the rebound. If the doors close her out, she gets mugged, is cared for by her unfaithful boyfriend, and

---

Ian S. Lustick, [ilustick@sas.upenn.edu](mailto:ilustick@sas.upenn.edu), Department of Political Science, University of Pennsylvania, 3440 Market St., Room 306, Philadelphia, PA 19104, is the author, *inter alia*, of *Trapped in the War on Terror* (University of Pennsylvania Press, 2006).

*Critical Review* 22(4): 427–447

© 2010 Critical Review Foundation

ISSN 0891-3811 print, 1933-8007 online

DOI: 10.1080/08913811.2010.541698

struggles within a web of deception about his infidelity, which leads to a crisis. In the language of counterfactuals there are at least two stories of Helen's future that are consistent with the laws governing the social, psychological, and natural worlds. Whichever narrative does occur will be seen, during and after its denouement, as the "factual" or "actual" world. The one (or ones) that do not occur would be termed "counterfactual." That a counterfactual outcome need not mean an impossible or even unlikely outcome is apparent from the fact that from the point of view taken at the moment of the sliding doors, looking forward into the future, each of these stories would be considered a "counterfactual"—a rendition of events that has not occurred, but which could occur, hypothetically at least, based on laws we believe in.

When seeking to learn from history, political scientists are forced to think in terms of counterfactuals: hypotheticals about what could have happened in the past but did not. An explanation of events usually contains implicit and explicit counterfactual predictions of what would have happened if the world were not governed by factors cited in the explanation. Highly plausible counterfactuals can render actual outcomes puzzling enough to require explanation in light of laws currently accepted as valid. Counterfactuals are also used to identify particular and previously overlooked circumstances capable of explaining, within the array of histories that might have occurred, the one that did occur.

In a series of works published between 1996 and 2006 Philip Tetlock and his collaborators studied counterfactual thinking by "experts"—mainly social scientists and historians—as part of a broader effort to understand the extent to which humans can operate as rational Bayesians, who continually update their theories based on new evidence, as opposed to being prisoners of preconceptions and of non-Bayesian heuristics. I will analyze five key works to construct a stylized account of Tetlock's evolving engagement with this issue.

### *Tetlock's Trajectory: From Optimism to Pessimism*

Tetlock's stance at the beginning of this period, documented in his co-edited 1996 volume with Aaron Belkin, *Counterfactual Thought Experiments in World Politics*, was strongly positivist and optimistic, and aspired to establish "best practices" for distinguishing sound from unsound counterfactuals. Their six rules for producing and interpreting counterfactuals in

the domain of world affairs were derived from the logical necessity of counterfactual thinking for any causal argument and were intended to make counterfactual analysis more scientifically rigorous.

Tetlock's 1999 *American Journal of Political Science* article on "Theory-Driven Reasoning" investigated the extent to which using counterfactuals to improve human performance in analytic and learning tasks might be compromised by entrenched styles of thinking—whether cast in ideological or cognitive terms.

Tetlock's 2001 article in the *American Political Science Review*, co-authored with Lebow, was distinctly pessimistic, emphasizing how rarely experts learn from information that is contrary to their political sympathies and how rarely they resist distortions in the processing of counterfactuals arising from their established cognitive styles.

In Tetlock's 2005 masterwork, *Expert Political Judgment*, he shifted primary focus to an even more basic impediment to the practice of Bayesian rationality. While the volume considers methodological, ideological, as well as cognitive factors, it attends most importantly to psychological factors under the broad rubric of "prospect theory"—treating human "psychologic" as including heuristics that interfere in perhaps irremediable ways with exploitation of counterfactual thinking via Bayesian logic.

Finally, *Unmaking the West* (2006) is an edited volume of essays by historians who offer and evaluate counterfactuals bearing on the question of whether we should view the actual, historical rise of the West as possible but *a priori* implausible, plausible but *a priori* improbable, probable but *a priori* not inevitable, or inevitable. In Tetlock's (co-authored) essays discussing the studies reported in the volume, his frustrations with counterfactual thinking—arising from methodological, ontological, cognitive, political, and psychological sources—are so great that he seems almost ready to throw up his hands at the project of substantially improving experts' use of information that runs contrary to their preconceptions, even while acknowledging the risk of occupying a position that he senses is uncomfortably close to the epistemological abyss.

In the trajectory of the specific questions posed in these works and in the results reported, one can detect a somewhat painful process of learning about how difficult it is to conduct counterfactual analysis rigorously. The decade of work begins with a methodological focus, with Tetlock and his collaborators seeking to improve social-science

practice by developing sensible rules of thumb for learning from counterfactuals in the face of the “ontological” reality that only one historical path can actually be observed. In other words, the only control groups we have for the history of a large social or political formation are comprised of elements produced by human imagination. But the contributors to the Tetlock and Belkin volume were far from convinced by the argument advanced by the editors. Perhaps in response to the arguments and disagreements provoked in this volume, Tetlock’s subsequent work has examined the obstacles experts face in producing sensible, scientifically grounded methods, beginning with political or ideological impediments and moving on to cognitive and psychological barriers.

Throughout this body of work Tetlock has brilliantly utilized experimental techniques to advance our understanding of the limits of the “*Homo scientificus*” model of human beings, including expert human beings. However, I will also suggest that theory-driven experimentation with computer-assisted agent-based models presents exciting opportunities to satisfy some of Tetlock’s most serious concerns. This is an approach acknowledged in his 1996 volume, but its implications and significance have gone largely unnoticed in his larger research program.

### *Criteria for Methodological Rigor*

In 1996 Tetlock and Belkin explored difficulties that counterfactual methodology pose for a variety of approaches to world politics: idiographic/discursive, statistical, rationalist, and computational. Any research approach needs clarity, parsimony, and theoretical explicitness, but counterfactual analysis requires additional criteria, particularly *cotenability* and *projectability*.

Cotenability, Tetlock and Belkin (1996b, 18) explained, is the requirement to “specify connecting principles that link the antecedent [what is imagined to have been different in the counterfactual] with the consequent [the putative result of the change in the antecedent]. . . that are cotenable [consistent] with each other and with the antecedent.” Projectability, offered as a distinct criterion, is closely related to cotenability. A finding, or a claim, is projectable if it applies beyond a single case; theoretical work is limited if it bears on only one specific outcome. Projectability is the “ability to predict what will happen in

new, hitherto unobserved cases. The same causal principles that allow us to retrodict the past should allow us to predict the future” (ibid., 30–31).<sup>1</sup> Citing Nelson Goodman’s work, Tetlock and Belkin (ibid.) describe a theory that results in claims that have projectability as the “acid test of scientific legitimacy.” In other words, projectability means that accurate claims or predictions arising from counterfactual analysis must be advanced as claims about *patterns*, not points, so that they are cotenable, as it were, with patterns of historical or future outcomes.

Most of the contributors to the Tetlock and Belkin volume respond in one of three ways to the challenge of cotenability in counterfactual analysis. Some contributors, especially James Fearon, Ned Lebow and Janice Stein, and Robert Jervis, find this criterion so daunting that they urge political scientists to abandon, or to impose severe limitations on, counterfactual analysis. These analysts emphasize the intractable difficulty of knowing, beyond very immediate consequences and causes, the relationship of counterfactual antecedents to consequents, as well as the difficulty of selecting, out of the contextual or background antecedents, which antecedent would have had to have been different for the counterfactual consequent to have occurred.<sup>2</sup> On the other side of the argument, Edgar Kiser and Margaret Levi, Barry Weingast, and Steven Weber, who endorse counterfactual thinking for different purposes either by judging as less daunting the problems establishing cotenability that Fearon, Lebow and Stein, and Jervis emphasize, as per Kiser and Levi and Weingast, or, as per Weber, despite the difficulties involved.

A third group of contributors, representing discursive-idiographic, game-theoretic, and computational approaches, accept the cotenability criterion as a standard, believe their techniques of counterfactual analysis abide by this standard, and (at least partly as a consequence) produce useful results. These contributors include George Breslauer, Yuen Foong Khong, Lars-Erik Cederman, Mark Turner, and Tetlock and Belkin themselves.

Cederman’s computer simulation agent-based modeling chapter embraced the methodological guidance offered by Tetlock and Belkin. The significance of his contribution, along with the implications of the failure of the editors to appreciate it, will be discussed below. What I note here is that the large majority of the contributors to the 1996 volume actually *rejected* the overall position of the editors, a position based on two premises: (1) The social world is complex but not so densely populated by chaotic relationships (butterfly effects) as to make

causal theories engaging that complexity impossible; and (2) workable rules for generating and evaluating counterfactuals should play a crucial role in the search for explanations and reliable forecasts.

Fearon, Lebow, Stein, and Jervis rejected the first proposition. They portrayed a political world so informed by complexity and “strong chaos” effects that projections of causal chains inevitably fail to correspond to outcomes. Accordingly, they also rejected the second proposition, which makes sense only if the social world, albeit complex, is imagined as analytically tractable. In other words, since theoretical social science cannot successfully treat the complexity of the real world, then there is no basis for improving our treatment of counterfactuals by sharpening our methods. On the other hand, Edgar Kiser and Margaret Levi, Barry Weingast, and Steven Weber also reject the first proposition, but embrace counterfactual reasoning either by substituting a “simple” game-theoretic world for the actual world, in which case the second proposition can be accepted (Kiser, Levi, Weingast); or by rejecting the second proposition in favor of purely speculative and “thought expanding” exercises in scenario-building (Weber).<sup>3</sup>

### *Ideological Thinking and Counterfactuals*

In light of the reaction of most of the volume’s contributors to the methodological approach to counterfactuals that formed the basis of the work, it is understandable that in followup research, including the 1999 article I consider next, Tetlock shifted his focus. Instead of seeking to improve experts’ use of counterfactuals for understanding the social world, much of his subsequent work in this area has focused on understanding how experts behave and how they respond to counterfactual reasoning and counterfactual information. The driving question in these studies is to identify the conditions, one might say pathologies, that interfere with the ability of thoughtful people, including experts in the domain of world affairs, from behaving according to Bayesian norms for rational updating. Abiding by these norms is crucial for the *Homo scientificus* model of man that underlies the recommendations for systematic use of counterfactuals offered by Tetlock and Belkin in 1996.

In “Theory Based Reasoning,” Tetlock sought to test the “cognitivist consensus” that heuristics trump Bayesian rationality by taking experimentation out of the laboratory, where highly charged prompts might

amplify the effect of heuristic biases on impressionable college sophomores. Instead he investigated the behavior of trained experts in response to counterfactual histories and to their own production of counterfactuals. In distinct experiments, Tetlock asked 52 specialists on the former Soviet Union and 25 experts on South Africa to rate the plausibility of counterfactual alternatives to actual outcomes at key points in the histories of those two countries. In a separate experiment, domain experts were asked to revisit and discuss predictions they had made five years before that had turned out to be *incorrect*. In other words, these experts were asked to comment on their own counterfactuals.

Overall his findings are typical of those produced within the prospect-theory research program: that is, that experts use simple rules of thumb as well as ingrained heuristics; they are inclined to protect themselves against loss of face rather than learning from errors; and they use politically grounded shortcuts instead of Bayesian norms. Indeed, it is striking how resistant experts were to the kind of Bayesianism implicit in Tetlock and Belkin's methodological advice for improving counterfactual thinking. Although the experiments did not show that the deep structure of thinking styles by "conservatives" versus "liberals" had any effect, they strongly suggested that concrete political sympathies led experts to dismiss counterfactuals that ran contrary to their sympathies, even as they embraced counterfactuals, or protected them from critical scrutiny, if they were in line with pre-existing political sympathies.

Tellingly, Tetlock also found that even in the midst of rejecting counterfactuals as unrealistic or immaterial, experts regularly resort to second-order counterfactuals of their own making to see the historical narrative as returning to the track they had predicted or that they preferred. This finding has potent implications for the positions of those authors in the Tetlock and Belkin volume who rejected counterfactual reasoning. It also reinforced Tetlock's shift of focus from methodological rules of thumb for improving counterfactual thinking toward understanding the way that thinking styles, motivated by ideological or theoretical predilections, shape expert use of counterfactuals regardless of methodological warrants or commitments. His most nuanced finding in this paper pertains to a dichotomous thinking-style variable established as a dependent variable in a study published in 1998. In that paper Tetlock corroborated work by others that found a tendency for

theory-driven forms of thinking . . . [to] dominate judgments of antecedent-consequent linkages (assuming *x* took this form, then *y* would have occurred); [while] more data-driven forms of thinking . . . dominate judgments of the plausibility of supposing that antecedent historical conditions could have taken on forms different from those they took on in this world. (Tetlock 1999, 338, citing Tetlock 1998)

In his 1999 article Tetlock characterizes “the key question” in the first study reported as not to discover how to profit more from counterfactual thinking, but to know “the extent to which counterfactual reasoning about major historical events is theory-driven (predictable from abstract preconceptions) as opposed to data-driven (constrained by the peculiar “fact situation” of each historical episode)” (Tetlock 1999, 336). Thus “cognitive style,” rather than appropriate method, came to occupy the center of Tetlock’s analytic attention. He found that experts are inclined to apply theoretical reasoning to generate or evaluate counterfactuals about the consequences of hypothetically stipulated differences in antecedent conditions. However, when asked to evaluate the counterfactually distinct antecedent conditions themselves, they appeal to the peculiar facts of the case rather than to overarching theory.

This fundamental distinction—between nomothetically animated thinking styles and idiographically animated thinking styles—emerged as the foundation for much of Tetlock’s subsequent work in this area.

In 2001, Tetlock and co-author Ned Lebow (2001, 842) relabeled this distinction with at least four synonymous pairs of terms for “modes of information processing”: “generalizers” vs. “particularizers,” “hedgehogs” vs. “foxes,” “close-mindedness vs. open-mindedness,” and “theory-driven” vs. “imagination-driven.” The master question guiding the study was, “How can we avoid becoming prisoners of our misconceptions?” (*ibid.*, 834). The hope—the hypothesis—was that encouraging, if not forcing, experts to engage in counterfactual thinking would diminish their disinclination to learn from errors or to avoid gross overcommitments to unwarranted beliefs. We see in this approach more evidence of the shift in Tetlock’s thinking about counterfactuals from the work done with Belkin in 1996. Apparently discouraged from promoting methodological best practices for learning logically and systematically from counterfactual analysis, his objective here was to begin with the assumption that experts depart from Bayesian rationality, and then to use counterfactual exercises to identify patterns in these departures and, perhaps, learn how such

exercises might mitigate propensities to err associated with non-Bayesian heuristics and psychologies.

Tetlock and Lebow's 2001 paper reports the results of a series of experiments in which the authors expose military historians and political scientists specializing in international politics and war to counterfactual accounts of important and familiar events, such as the Cuban Missile Crisis, the outbreak of World War I, the defeat of Nazi Germany, the origins of the Cold War, and the "rise of the West." Prior to these exercises, they categorized the experts according to how strongly they viewed standard "neo-realist" or "balance of power" theories as master accounts of international relations, and with respect to their "need for explanatory closure." Subjects in treatment groups were engaged in moderate or intensive consideration of counterfactual histories (for example, trajectories of the Cuban Missile Crisis that did not occur but that might well be imagined as having occurred). Subjects were then compared with respect to how they responded to these new counterfactuals, and with respect to how exposure to those counterfactuals affected their own subsequent characterization of the rate with which actual outcomes became inevitable and alternative outcomes became impossible.

The pattern of the findings in this study is complex and in some respects inconclusive. The authors report that experts with pre-existing cognitive commitments to "covering-law" explanations (where experts believe that certain abstract principles provide consistent and satisfying causal accounts of international outcomes), as well as experts who seem relatively intolerant of ambiguity and strive hard for cognitive closure, were more resistant to counterfactuals than those who did not score highly on these two measures. They also emphasize that, across the board, engagement with counterfactuals led experts to increase their assessment of the probability of events that did not actually occur, even to the extent of producing a sum of probabilities greater than 100 percent. The most potent finding, however, was that as salutary as counterfactual thinking may be for opening the minds of experts drawn prematurely to firm conclusions, it was as dangerous for the open-minded because of the seductiveness with which the vivid but irrelevant and logically uninformative details involved in those accounts encouraged unwarranted commitments to their plausibility or even probability.<sup>4</sup> These two counterposed findings led the authors to cast doubt on the commitment to the positivist or neo-positivist methodology espoused in

the Tetlock and Belkin volume, concluding this study on a skepticist note, *viz.* that there is an “inescapable trade-off between being close-minded but logically coherent or open-minded but logically incoherent” (Tetlock and Lebow 2001, 830).<sup>5</sup>

The advice offered by Tetlock and Lebow (2001, 843) to policy-makers and analysts, and to those pondering their putative wisdom about alternative/counterfactual futures, is not to rely on good theory, or good methods, but on “good judgment.” The echo of Aristotle here should not be considered coincidental, since in that tradition “good judgment” is a capacity not reducible to laws or valid theories. In that context it is also no coincidence that Tetlock’s next and by far most influential contribution to the literature on counterfactual thinking was his 2005 multiple-prize-winning book, *Expert Political Judgment (EPJ)*.

### *Expert Political Judgment*

Tetlock’s explicit purpose in *EPJ* is to incentivize pundits, and other prognosticators who have the attention of the public, to acquire the “good judgment” necessary to distinguish reasonable from unreasonable counterfactuals, and to distinguish reasonable from unreasonable inferences from counterfactual thought experiments. Sewn together in a volume of extraordinary power and depth, the separate studies produced by Tetlock’s research program establish the framework for posing the problem in this way. Significantly, the chapters exploit, expand, and gloss the results of work on counterfactuals by Tetlock and his collaborators *since* the 1996 Tetlock and Belkin volume. In other words, in *EPJ* the challenge is *not* to identify “best practices” for establishing “sound counterfactuals,” but to reduce, without eliminating, expectations that scientific norms will improve the production and evaluation of public policy.

The main argument of the book can be summarized as follows. Expertise matters very little for improving the accuracy of predictions by particular experts who, by many key measures, hardly do better than dart-throwing chimpanzees. By contrast, statistical models and extrapolation techniques from historical baselines, neither of which entail substantive understanding or theory, do much better at this task than the chimps—and the experts. On the other hand, averages of expert predictions do dominate individual expert forecasting accuracy. Tetlock

also finds that experts and decision makers are generally so interested in finding true positives that relatively high rates of false positives and false negatives are deemed acceptable. This provides a basis for continued interest in how to harness experts' knowledge by holding them accountable and encouraging them to think in disciplined counterfactual terms. The experts' success rates and reactions to the failure of their predictions show that cognitive style ("hedgehogs" vs. "foxes") matters a great deal. On almost all measures foxes outperform hedgehogs, not only in terms of the accuracy of forecasts but in the extent to which they use counterfactuals to learn from error.<sup>6</sup>

In *EPJ* Tetlock's analysis of the sources of poor expert performance extends beyond political sympathies, ideological postures, or cognitive styles.<sup>7</sup> From the beginning of his consideration of counterfactual reasoning, Tetlock has recognized the intrinsically difficult problem of learning about general laws from a single set of historical events. In *EPJ* he refers to this problem as stemming from "the ontological inadequacies of history as a teacher" (Tetlock 2005, 161). If in the 1996 volume he emphasized methodological strategies for becoming better students of an admittedly poor teacher, in *EPJ* he discounted the importance of such strategies in view of "our psychological inadequacies as pupils" and wondered if it "begins to look impossible to learn anything from history that we were not previously predisposed to learn" (*ibid.*). The pathologies he identifies as accounting for most of the disappointing behavior of experts are deeper aspects of human psychology than his earlier work had emphasized: human psychologies that produce or accentuate methodological errors or that favor certain (distortive) heuristics over others.

This line of analysis is evident in Tetlock's explanation of why hedgehogs score lower on most of his metrics than do foxes. It is not because of an intrinsic error associated with attachment to an *ex ante* theory (which after all is a legitimate aspect of any neo-positivist science), but to the greater vulnerability of hedgehogs to three well-documented human tendencies that distort (Bayesian) processing of information: hindsight bias, the fundamental attribution error, and cognitive conservatism.<sup>8</sup> Hindsight bias is the powerful inclination for humans to treat what they believe to have actually happened as more probable to have occurred than it was prior to the fact. Hindsight bias tends to undermine the credibility of even the most reasonable and minimalist of counterfactuals. While Tetlock (2005, 37) argues that counterfactual exercises

can “check our susceptibility to hindsight bias,” the salutary effects of these exercises have relatively little impact on hedgehogs compared to foxes because of the rapidity and firmness with which hedgehogs assimilate information in tension with their priors, as if it were consistent with them.

People manage to convince themselves, sometimes within milliseconds, that they knew it all along. This explanation [that unconscious and immediate cognitive processing tends to render observed outcomes consistent with what it was that was thought would occur] dovetails nicely with the greater propensity of hedgehogs to exhibit [hindsight bias]. Hedgehogs should place a higher value on cognitive continuity, on minimizing gaps between their current and past opinions. Hedgehogs should thus be more predisposed—by dint of their cognitive and emotional makeup—to assimilate outcomes, as soon as they become known, in terms of their favorite explanatory categories. (Ibid., 139–40)

The fundamental attribution error entails assigning undesired outcomes (including erroneous predictions) to exogenous factors (unusual “task difficulty, unfair testing conditions, or bad luck”) while attributing desired outcomes (including correct forecasts) to the validity of one’s own theory or analysis or the “shrewdness of one’s opinions” (Tetlock 2005, 128). Among the experts studied by Tetlock, hedgehogs indulge in this behavior significantly more extravagantly than foxes. Hedgehogs also display a more intense “cognitive conservatism” than foxes; indeed, it is almost a defining characteristic of hedgehogs. Cognitive conservatism refers to what Tetlock (ibid.) calls the basic “reluctance of human beings to admit mistakes and update beliefs.” Cognitive conservatism exerts such a strong influence on hedgehogs that, although they are otherwise resistant to counterfactual thinking because of its tendency to highlight the arbitrariness or contingency of accurate forecasts, hedgehogs *embrace* counterfactuals, even unsound ones, if doing so can help protect preconceptions from contrary data. In this way “counterfactual history,” writes Tetlock (ibid., 131), “becomes a convenient graveyard for burying embarrassing conditional forecasts.”

In *EPJ*’s final chapter Tetlock attends explicitly to the epistemological implications of the pattern of his findings. Since history cannot be repeated, counterfactuals are necessary for causal inference and accurate forecasting. The fact that even experts are susceptible to crippling psychologies makes it difficult or even impossible to distinguish sound

from unsound counterfactuals or to responsibly evaluate *ex ante* beliefs about causality or causal inferences. Tetlock measures four epistemological positions against these conclusions. His technique for doing so is to present a dialogue among the four: the hardline neopositivist, the moderate neopositivist, the reasonable relativist, and the unrelenting relativist. Instructively, it is only the unrelenting relativist who discusses the goal of the book—enhancing capacities for “good judgment”—but his conclusion is that such a goal is itself a snare and a delusion (Tetlock 2005, 221 and 227).

At the end of the book Tetlock (2005, 229) acknowledges uncertainty as to whether he is an “epistemologically liberal neopositivist” or an “epistemologically conservative relativist.” Indeed, the only evidence that Tetlock still identifies with the methodological goals set out in his 1996 study of counterfactuals in world politics is the rigor of the methods he uses and reports in *EPJ*, and his own apparent ability to change his mind as a result of his research.

### *Unmaking the West*

The last of Tetlock’s contributions to research on counterfactuals to be considered in this essay is the volume *Unmaking the West: “What-If?” Scenarios that Rewrite World History* (2006), which Tetlock edited with Richard Ned Lebow and Geoffrey Parker. The book includes introductory and concluding chapters by Tetlock and Parker. Although this work was published most recently, it is apparent from references in the articles and books discussed above that Tetlock and his collaborators were in contact with at least a large subset of the contributors for many years and have used their thoughts on the rise of the West (contingent or inevitable?) to inform much of their thinking.

In this volume Tetlock continues to retreat from the project of improving social science by designing best practices for the identification and use of sound counterfactuals. In some respects it represents a substantial abandonment of key elements in the original research program.

The volume is structured as an opportunity for historians with different points of view and substantive specializations to produce accounts of historical trajectories that differ from the actual trajectory that (it is believed) human beings have traced in recent centuries. The

key aspect of that trajectory is figured as the “rise of the West” as the technologically, economically, politically, and militarily dominant region in the world, and the extension of that supremacy into the contemporary period. The purpose of the volume is to challenge historians, such as E. H. Carr, who imagined their task as identifying “why events had to take the path they did [and] to explain why *ex post facto* no other path was possible” (Tetlock and Parker 2006b, 30). In their introduction, Tetlock and Parker explicitly characterize Carr’s conception of the historian’s task as a flagrant example of “hindsight bias.” Indeed the entire volume is imagined as an exercise in combating hindsight bias. Drawing on the basic tropes of prospect theory, the editors encourage their contributors—trained historians—to shift their definition of historical problems. Instead of framing a question in a typical, Carr-like manner, with a “factual frame,” each contributor was challenged to frame his or her question about the rise or sustained dominance of the West counterfactually. “Factual framing” asks why a particular event occurred. Framing the question “counterfactually” asks why alternatives failed to occur (*ibid.*, 23).

Note the absence here of any focus on how to bring the behavior of historians or other experts more in line with Bayesian norms of error correction and learning. Nor, as is apparent from the concluding chapter by Tetlock and Parker, do the authors imagine that their elaborate exercise has actually led to better answers to the question of why the West rose, and why it maintained its supremacy, than were available before. Rather, the success they identify in the volume is to have replaced images of the rise of the West as a single narrative. Instead of history as a series of episodes flowing in perfectly understandable if not inevitable fashion from preceding circumstances, they claim, almost tritely, that the combined effect of the chapters in this volume

revealed the extreme complexity of the process we call the rise of the West. It involved the coming together of a large number of changes and processes in Western and global history that jointly produced the modern world . . . [making] us appreciate the multitude of paths that led into this world. (Tetlock and Parker 2006c, 389)

As Tetlock’s objective in EPJ shifted from achieving better understanding of the world via counterfactual reasoning to better understanding of experts; so, too, in this volume his focus shifts heavily: away from interest in the substantive question of the rise of the West and

toward the behavior of historians and the extent to which counterfactual exercises weaken the pathological effects of hindsight bias. In *EPJ* Tetlock concluded, as noted above, by exhorting experts to develop “good judgment” without explaining a method for doing so. In *UTW*, Tetlock and his collaborators offer an even weaker conclusion. There is no exhortation to historians to develop the “good judgment” that would allow them to distinguish which factual or counterfactual frames to use in a given circumstance; there is only a plea to be aware that the history they believe as actual was not inevitably so (Tetlock and Parker 2006c, 366, 389).

Many of the contributions in this volume are fascinating and erudite. A few chapters reveal the authors to command a sophisticated sense of the implications of thinking about historical outcomes with the distributional and probabilistic perspective appropriately associated with thinking about the future.<sup>9</sup> But some contributions adhere so powerfully to the political and analytic preconceptions of the authors to buttress judgments that history “had” to be the way it was, eventually; or that a favorite person, event, or contingency was entirely responsible for all that subsequently occurred; as to corroborate Tetlock’s findings that psychologic usually trumps methodology.<sup>10</sup> Fortunately, the collection includes valuable comments by historians that correct or at least counterbalance the extravagant tendencies of some of their colleagues to exercise their imagination brilliantly, but not in close conformity to the gentle methodological guidance offered by the editors.<sup>11</sup>

### *Saving Tetlock’s Ambition from His Findings*

In the body of work covered in this essay, Tetlock regularly invokes the point made most memorably by Stephen J. Gould: that if only the “tape of history” could be replayed with slight changes of our own choosing, then vast opportunities for testing counterfactual claims about possible, plausible, and probable worlds would be opened up (Tetlock and Belkin 1996a, 9; Tetlock 1999, 337, 343, 359; Tetlock and Lebow 2001, 840; Tetlock 2005, 146; Tetlock, Lebow, and Parker 2006a, 3, 18, 283). Instead, however, students of all historical sciences, including the social sciences, are stuck with just one thread of actuality (however variously it may be recorded and remembered).

The fact that we have only one actual or observable history is what Tetlock (1999, 337) refers to as the ontological basis for his judgment that however poorly we behave as students, “history is a terrible teacher.” The thrust of his published work on this problem between 1999 and 2006 has been to maintain that in the face of this “ontological” obstacle, and considering strong human psychologies that contradict Bayesian rationality requirements, the elaborate guidelines promulgated in the 1996 volume with Belkin must be shelved. Instead of devising more efficient rules for distinguishing sound from unsound counterfactuals, Tetlock is satisfied (if grudgingly so) to praise “good judgment” (absent advice for how that talent may be acquired) to encourage experts to appreciate that the world as it is is simply *not* the only way it could have been, and to warn us to beware the seductive distortions of vivid and detailed stories.

Yet Tetlock should not feel compelled to retreat so completely from his 1996 ambitions. For there is a crucially important sense in which the tape of history *can* be rerun. The requirements for this task are now available: increasingly powerful social-science theories for constructing models of important domains, and easily accessible computer simulation programs for deploying those theories in fully transparent possible worlds.<sup>12</sup> By stochastically perturbing dynamic “runs” of theoretically specified models, batches of possible worlds can be produced. Each trajectory within such a batch is automatically and fully “cotenable” (to invoke the single most important methodological rule offered by Tetlock for evaluating counterfactuals) with the operationalized theories used to build the model. Standard statistical operations can then be performed on the populations of trajectories within one or more distributions generated by experimentally coherent variations in theoretical claims, weights of variables, parameter settings, or initial conditions.

The fact is that in 1999 Tetlock himself imagines this possibility, at least in principle. In that article he extends Gould’s thought experiment about rerunning the tape of history to show how the methodological guidelines he had developed with Belkin in 1996 could be obviated by social scientists seeking well-grounded and systematic opportunities to evaluate counterfactuals. To do so, Tetlock draws explicitly on theories of modal logic developed by philosophers of science. They depict an actual world surrounded by large swarms of “accessible” but not actual “possible worlds.”<sup>13</sup> Using the theory of possible worlds, Tetlock imagines that counterfactuals in the form of forecasts could be evaluated

as “‘frequentist’ claims about the distribution of possible worlds in repeated-simulation reruns of history” (Tetlock 1999, 360). This stance opens up the distinct possibility that a prediction about the future could be correct, in the sense that most of the possible outcomes are consistent with it, even though in the “actual” future the predicted state of affairs does not materialize.

Tetlock (1999, 360) imagines this procedure but then dismisses it for essentially rhetorical reasons—“People,” he says, “will tire of hearing that the world they happen to inhabit is vanishingly improbable.” What is a bit strange is that one of the most exciting contributions in the Tetlock and Belkin volume was Lars-Erik Cederman’s chapter “Rerunning History by Computer.” In that essay Cederman used just this technique of generating distributions of trajectories out of treatment and control conditions to test conflicting hypotheses about world politics.<sup>14</sup> Specifically, he reported on research asking whether the ultra-stability of the European power politics system (defined as regular and sustained competition among a plurality, but not a very large plurality, of states) should be considered as an automatic outcome of anarchic competition, or whether the actual history we have observed was only one type that we might have observed. Cederman’s essay is an exemplar of ensemble-type agent-based modeling research—research focused on worlds that are stylized for purposes of focusing on specific questions posed by contending theories, much as a laboratory “stylizes” the world to achieve an artificially but legibly simple world for evaluating theoretical claims.

Arguably, of all the chapters in the Tetlock and Belkin volume, Cederman’s essay was the most consistent with the strictures for generating and evaluating counterfactuals advanced by the editors. Yet while it was referenced a dozen times throughout the volume, neither the editors nor the contributors responded to the opportunity Cederman’s technique presented for addressing the deep methodological issues raised.

To be sure, computer-assisted agent-based modeling simulation, as a method for conducting social-science theory-driven research, is in an “adolescent” stage. It is no longer in its infancy, but it is still uneven in its accomplishments, uncertain of its norms and capabilities, and more apt to exaggerate or underestimate its potential than to offer a mature, balanced assessment. Although numerous “abstraction” studies are available that leverage the computer’s ability to display the results of immense numbers of very simple interactions, as are a fair number of excellent “ensemble”

studies—using substantive theory to analyze generic problems in a virtual laboratory—there is much to be done before multiple theories can be used to “virtualize” specific problems of interest to scholars and policy makers.<sup>15</sup> But its potential is undeniable, especially for the kind of problems facing Tetlock and others interested in grounding social science in neo-positivist (by which I mean Lakatosian) methods.<sup>16</sup>

In the body of work reviewed in this essay it is easy to see the role that even simple applications of this technology could play. Impossibility claims can be challenged by showing futures arising from clear and limited assumptions. Futures can be treated as close to impossible if they cannot be observed, despite implementation of wide variation in parameter settings and theoretical assumptions. Cognitive conservatism can be mitigated by illustrating how much variation can be generated out of the same basic theoretical position. Disputes over the decisiveness of particular circumstances can be refereed by evaluating the ease with which comparable futures can be generated by varying those circumstances. Possibility distributions generated from well-specified models can act as pruning shears, yielding hypotheses to distinguish the plausible from the possible and the probable from the plausible.

Philip Tetlock’s work has contributed brilliantly to our understanding both of the importance of disciplined thinking about counterfactuals, and the difficulties of doing so. The argument offered here is that although Tetlock has shown how rare worlds are in which counterfactuals are generated and used responsibly, that kind of world may be, in fact, much more accessible than his recent work would lead us to believe.

## NOTES

1. The same basic idea was advanced by Imre Lakatos (1970, 134) in his insistence on the “prediction of new facts” as a crucial arbiter of good science.
2. Similar stances are taken by Bernhard Kittel (2005) and James Mahoney (2005).
3. For a detailed and critical analysis of the methodological approaches taken by each of the contributors to the 1996 Tetlock and Belkin volume see Lustick 2006.
4. This is the “black swan” effect. See Nassim N. Taleb (2007). For this effect Tetlock and Lebow cite Koehler 1991 and Tversky and Kahneman 1983.
5. The very last paragraph of the article adds considerably to the warrant for interpreting the authors’ conclusions in this manner. In this passage the authors contrast “positivism” (never distinguished from “science”) with “constructivism,” suggesting (oddly, in my view) that the latter is not or cannot itself be “positivist,” and at the same time promoting “constructivism” (stripped of any

- claim to objective theoretical content) as equally worthy of expert and scholarly attention (Tetlock and Lebow 2001, 843).
6. The one area in which hedgehogs out-perform foxes is in their greater resistance to warrantless commitments to counterfactuals rendered seductively vivid by black swans (Tetlock 2005, 164–88).
  7. Performance is imagined in *Expert Political Judgment* as measured by the standards of *Homo scientificus*.
  8. Each of these psychologies appear in previous work by Tetlock on counterfactuals, but only in *Expert Political Judgment* do they begin to carry much, if not most, of the explanatory burden.
  9. In his contribution to *Unmaking the West* Joel Mokyr (2006, 277–322) explains the West's rise probabilistically on the basis of the early development of an abstract theory-oriented scientific stance without ruling out Chinese industrialization primacy despite the absence of this dominant metaphysic in their philosophical tradition. In another chapter Holger H. Herwig (2006, 323–60) responsibly exploits Black Swan techniques of vivid evocation of counterfactuals in service of a second order counterfactual to attain the minimalist objective of combating hindsight bias. His argument is that there was not only one way Hitler could have been defeated.
  10. In his energetic chapter on Themistocles, Victor Davis Hanson (2006, 47–89) embraces the great-man theory, conservative political principles, and a host of convenient but implicit and unexamined theories to hang most of world history on one man's stratagem in the battle of Salamis. Carlos M. N. Eire (2006, 119–42) combines rabbit-out-of-the-hat theoretical claims with a series of analogical assertions and "common sense" claims to attract support for a spectacular Black Swan scenario—Jesus is not crucified, leading Constantine to emerge as a Jewish Caesaro-Papist Emperor of a united Roman Europe.
  11. In this regard see especially chapters in *Unmaking the West* by Barry Strauss (2006, 90–118), and Carla Gardina Pestana (2006, 197–202).
  12. To be sure, these worlds are "virtual," but it is profoundly important in this context to keep in mind that *all* worlds constructed by scientists out of their theories, models and assumptions, and whether in laboratories or not, are "virtual" realizations that are stylized, demarcated, and isolated enough from the "real world" to be systematically studied.
  13. For a useful collection of opposing views by philosophers about how to think coherently about "possible worlds" that are not actual, but do "exist," see Michael J. Loux (1979). While this work is an excellent introduction to the dominant concerns of philosophers with respect to the terminological challenges of using "actual" to mean something other than "existing," the approach of these scholars could itself have been greatly assisted had they imagined the use of computer programs to generate virtual trajectories that do "exist" but are not "actual."
  14. Tetlock and Belkin (1996b, 16) more or less dismiss Cederman's work by commenting that "a counterfactual grounded in an elegant computer simulation might blow a gaping hole in an influential theoretical argument but tell us precious little about the actual world it supposedly simulates." This comment

assumes what Tetlock elsewhere (where he explicitly espouses a Lakatosian epistemological perspective) shows he understands to be impossible, i.e. a non-theory laden understanding of “the actual world.” James Fearon’s essay in this volume does discuss cellular automata (a form of agent-based modeling) but suffers from grievous confusions about the technique, including conflation of ambitions for point predictions with exploitation of variation in distributions of outcomes for probability forecasts. See Lustick 2006.

15. See Lustick and Miodownik 2009.

16. For Tetlock’s own Lakatosianism see Tetlock 2005 (179n) and Tetlock 1999 (349).

## REFERENCES

- Eire, Carlos M. N. 2006. “The Quest for a Counterfactual Jesus: Imagining the West Without the Cross.” In Tetlock, Lebow, and Parker 2006a.
- Hanson, Victor Davis. 2006. “A Stillborn West? Themistocles at Salamis, 480 BC.” In Tetlock, Lebow, and Parker 2006a.
- Herwig, Holger H. 2006. “Hitler Wins in the East but Germany Still Loses World War II.” In Tetlock, Lebow, and Parker 2006a.
- Kittel, Bernhard. 2005. “The American Political Methodology Debate: Where is the Battlefield?” *Qualitative Methods* 3(1): 12–19.
- Koehler, Derek. 1991. “Explanation, Imagination, and Confidence in Judgment.” *Psychological Bulletin* 110(3): 499–519.
- Lakatos, Imre. 1970. “Falsification and the Methodology of Scientific Research Programmes.” In *Criticism and the Growth of Knowledge*, ed. Imre Lakatos and Alan Musgrave. Cambridge: Cambridge University Press.
- Loux, Michael J. 1979. *The Possible and the Actual: Readings in the Metaphysics of Modality*. Ithaca, N.Y.: Cornell University Press.
- Lustick, Ian. 2006. “Toward a Solution to the Problem of Counterfactual Reasoning: Exploiting Evolution, Complexity, and Computers.” Paper presented at the annual meeting of the American Political Science Association, 31 August. [http://www.allacademic.com/meta/p150928\\_index.html](http://www.allacademic.com/meta/p150928_index.html).
- Ian S. Lustick, and Dan Miodownik. 2009. “Abstractions, Ensembles, and Virtualizations: Simplicity and Complexity in Agent-Based Modeling.” *Comparative Politics* 41(2) (January): 223–44.
- Mahoney, James. 2005. “Clarifying Comparative–Historical Methodology.” *Qualitative Methods* 3(1): 19–22.
- Mokyr, Joel. 2006. “King Kong and Cold Fusion: Counterfactual Analysis and the History of Technology.” In Tetlock, Lebow, and Parker 2006a. 227–322
- Pestana, Carla Gardina. 2006. “Nineteenth-Century British Imperialism Undone by a Single Shell Fragment: A Response to Jack Goldstone’s ‘Europe’s Peculiar Path.’” In Tetlock, Lebow, and Parker 2006a.
- Strauss, Barry. 2006. “The Resilient West: Salamis Without Themistocles, Classical Greece Without Salamis, & the West Without Classical Greece.” In Tetlock, Lebow, and Parker 2006a.

- Taleb, Nassim N. 2007. *The Black Swan: The Impact of the Highly Improbable*. New York: Random House.
- Tetlock, Philip E., and Aaron Belkin, eds. 1996a. *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives*. Princeton: Princeton University Press.
- Tetlock, Philip E., and Aaron Belkin. 1996b. "Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives." In Tetlock and Belkin 1996a.
- Tetlock, Philip E. 1998. "Close-Call Counterfactuals and Belief System Defenses: 'I Was Not Wrong but I Was Almost Right.'" *Journal of Personality and Social Psychology* 75(3): 639–52.
- Tetlock, Philip E. 1999. "Theory-Driven Reasoning About Plausible Pasts and Probable Futures in World Politics: Are We Prisoners of Our Preconceptions?" *American Journal of Political Science* 43(2) (April): 335–66.
- Tetlock, Philip E., and Richard Ned Lebow. 2001. "Poking Counterfactual Holes in Covering Laws: Cognitive Styles and Historical Reasoning." *American Political Science Review* 95(4): 829–43
- Tetlock, Philip E. 2005. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton: Princeton University Press.
- Tetlock, Philip E., Richard Ned Lebow, and Geoffrey Parker, eds. 2006a. *Unmaking the West: "What-If" Scenarios that Rewrite World History*. Ann Arbor: University of Michigan Press.
- Tetlock, Philip E., and Geoffrey Parker. 2006b. "Counterfactual Thought Experiments: Why We Can't Live Without Them and How We Must Learn to Live with Them." In Tetlock, Lebow, and Parker 2006a.
- Tetlock, Philip E., and Geoffrey Parker. 2006c. "Counterfactual History: Its Advocates, Its Critics." In Tetlock, Lebow, and Parker 2006a.
- Tversky, Amos, and Daniel Kahneman. 1983. "Extensional versus Intuitive Reason: The Conjunction Fallacy as Probability Judgment." *Psychological Review* 90(2): 292–315.