AHR HISTORY LAB

David James Gill, Marc Trachtenberg, Michael J. Gill, Philip E. Tetlock, Thomas K. Robb, Michael E. W. Varnum, Cendri A. Hutcherson, Igor Grossmann, and Zoe Trodd

Predicting the Past

Testing Expert Historical Judgement

Absences pervade the historical record. The loss or destruction of material, redaction of documents, silence of participants, data embargoes, and poor record keeping present inherent difficulties to any understanding of the past. Gaps in the historical record pose significant challenges but they can also provide valuable opportunities. Instead of avoiding all gaps, historians can test the accuracy of some inferences about the past by carefully outlining their assumptions and explicitly predicting what they believe to have occurred in the absence of evidence. Subsequent discoveries or declassifications can then be used to assess the accuracy of these hypothesized explanations and, in turn, help us to evaluate the quality of historians' thinking about the unknown past. Given enough examples, we can begin to learn more about how to make better predictions about the past or what we term "retrodictions."[1]

The American Historical Review, 2025, 130(4), 1615–1630
https://doi.org/10.1093/ahr/rhaf590
© The Author(s) 2025. Published by Oxford University Press on behalf of the American Historical Association. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Downloaded from https://academic.oup.com/ahr/article/130/4/1615/8404686 by guest on 12 February 2026

The importance of prediction has been well documented across the social sciences for decades.[2] Economists, political scientists, and psychologists regularly extol the virtues of making testable predictions for good scholarly practice and improved judgement. Although historians and historically minded scholars in the social sciences routinely rely on their own expert judgement when studying the past, few have considered making and testing explicit predictions when conducting historical research. Even fewer have considered the methodological demands and theoretical implications of retrodiction.[3]

This article considers why historians and social scientists should attempt to predict the past and explains how they can make and test retrodictions in a more rigorous manner. These conceptual and methodological insights are of considerable significance. Retrodiction is not an abstract or hypothetical subject. While some scholars may argue that accurate prediction of the past is irrelevant or impossible, many historians and social scientists regularly make inferences about undocumented history. Educated guesses occur across the discipline and beyond. There is nevertheless a tension between the widespread but often implicit employment of retrodiction, and the rarity with which it is deliberately considered and applied. Thinking more explicitly about retrodiction offers the chance to assess and refine historical interpretation. By understanding why some predictions about the past prove more or less accurate than others, we can think more clearly about how to manage empirical gaps, better adjudicate between competing interpretations, and gain greater confidence in our own retrodictions.

This article proceeds in three parts before concluding. First, we introduce the concept of retrodiction and then reflect on the rewards and risks involved in making predictions about the past. Second, we consider different ways of making and testing predictions about the past, giving special attention to the important role of retrodiction tournaments. Third, we draw on key insights from research in the social sciences to inform the practice of retrodiction and highlight future avenues of research.

Historical claims are built on evidence. From archival research and oral histories to archaeological investigations and satellite imaging, we have assembled much of our knowledge about the past from available data. Yet there are many gaps in the historical record and the pool of data available to experts is often very shallow. Experts who study the past rely on imperfect or incomplete information. In developing a broader interpretation of history, they may therefore need to speculate or "predict" the sort of picture that would emerge if or when more evidence ever became available. This attempt to predict the past—to engage in retrodiction—is thus one of the basic techniques that historians and social scientists use when they construct an interpretation of the past.

Downloaded from https://academic.oup.com/ahr/article/130/4/1615/8404686 by guest on 12 February 2026

Efforts to predict the past are commonplace. Scholars often make assumptions or reach conclusions about the past in the absence of direct evidence. Many historians and social scientists therefore already make retrodictions, albeit sometimes unknowingly or implicitly.[4] For example, scholars have debated the status of the United Kingdom's considerable war debts to the United States government following World War I. At least one historian has concluded that the war debt issue was effectively laid to rest in 1940 as a new and highly personal relationship emerged between Winston Churchill and Franklin Roosevelt.[5] In the absence of direct evidence, however, this conclusion represents an implicit prediction about what occurred in the past. Recent archival discoveries drawing on data from both sides of the Atlantic challenge this claim by showing that the United Kingdom's war debts to the United States persisted as a source of diplomatic tension in the decades that followed and remain unpaid and outstanding.[6]

Some historians and social scientists advancing arguments about the past may seek to avoid making any explicit claims about gaps in the historical record and limit themselves to presenting the evidence they have found. Yet those scholars may have missed a valuable opportunity to test their historical judgement. Historians and historically minded scholars in the social sciences should instead accept that their interpretations of the past are often based on incomplete data and speculate about the sort of picture that would emerge if more evidence became available. They could then test their understanding of events by providing specific details about what we would expect to see if that missing evidence ever became public. Retrodiction does not alter the underlying logic of an historical argument but instead seeks only to make it clearer and more testable. Thinking more explicitly about retrodiction can therefore improve historical practice and our knowledge of the past. If historians and social scientists explain how they reach their conclusions, they and other scholars can better assess the quality of those arguments if new and important evidence ever comes to light. By making and testing enough examples, we may be able to determine how far the lessons of retrodiction can be generalized to other periods or fields of study.

We accept that many scholars prefer to focus their attention on explanation rather than retrodiction. Analyzing what the available evidence says is a more familiar practice than speculating explicitly about what presently unavailable evidence might show.[7] Yet explanation and retrodiction are cognate concepts.[8] As an interpretation is developed—and historians and social scientists develop an interpretation by drawing on theories about how the world works—predictions are generated as a matter of course. "At any given point in an historical argument," one of the authors of this article has previously explained, "the historian can ask, given what was said up to that point, whether it would be possible to predict how things would develop. This provides a useful test of the *power* of the argument: a strong interpretation should have a certain predictive force."[9]

Downloaded from https://academic.oup.com/ahr/article/130/4/1615/8404686 by guest on 12 February 2026

Retrodiction also provides the historian with a way to assess the *validity* of the interpretation taking shape in their mind. As historians or social scientists proceed with a project, they could ask themselves what might turn up when they read new material—whether material that has already been released, but which they have not seen yet, or material that is released later—if that tentative interpretation is correct. Those retrodictions about what new sources would show might be borne out, in which case the historian would have more confidence in the validity of that interpretation. Alternatively, the new sources might not be in line with what had been expected, in which case the historian might have to rethink their emerging interpretation, and maybe even question the assumptions or arguments on which it was based.

We argue that it is important for scholars to consider and apply retrodiction consciously and thoughtfully, by explicitly stating what they would expect to see in as-yet-unexamined or unavailable sources. Historians and social scientists have advanced broader interpretations about what drove the collapse of the Soviet Union, for example, with a small number of scholars making educated guesses about the unknown past and what they would expect the archives to eventually reveal.[10] The absence of evidence therefore serves as a way in which to test and potentially bolster a specific historical interpretation.

Transforming explanations of past events into explicit, testable predictions enables historians and social scientists to shift from rationalizing events after they have occurred to formulating hypotheses in advance. Emphasizing *a priori* theorizing over *post-hoc* analysis can potentially lead to sharper, more robust historical narratives and help to strengthen the inferential power of interpretations or theories regarding causes of past events and historical trends.[11] Retrodiction can also help scholars avoid hindsight bias—namely, the common tendency to incorrectly believe in retrospect that one correctly guessed how events would unfold.[12] By making falsifiable claims, before new historical data comes to light, historians and social scientists can combat this tendency and increase accountability, thereby fostering scholarly rigor.[13]

Attempts to predict the past, however, do carry some risks. Retrodiction requires a degree of uncertainty about what has happened in the past. If we make educated guesses when the required evidence is unavailable to us, there is always the possibility that the missing data will never come to light. Assessments of historical judgement require the discovery of new, relevant evidence after a retrodiction has been made. Some predictions about the past, no matter how specific or well formulated, may be unverifiable. Crucial archival records have been destroyed in wars, for example, just as some individuals or organizations may have never written much down. Historians and social scientists should therefore try to focus their attention on areas where there is a reasonable possibility of testing their retrodictions.

New data about the past can be anticipated in some cases. Some archives use "timed releases," such as the National Archives in the

Downloaded from https://academic.oup.com/ahr/article/130/4/1615/8404686 by guest on 12 February 2026

United Kingdom, which embargoes many government documents for twenty years before releasing them to the public. We tend to know which government documents have been embargoed even if we do not know what they say. Not all government documents are released within twenty years, to be sure, but even in sensitive cases a release date is often agreed. Venturesome scholars can test their historical judgement by offering educated guesses about what these files contain and help us learn from their retrodictions by sharing how they reached these conclusions.

Moreover, new data continues to emerge at a remarkable pace. Archives around the world regularly release new material. Previously closed archives have reopened, even if only temporarily, and archivists remain hard at work recovering lost data.[14] Archival data represents one of many types of evidence that can emerge over time. Witnesses may eventually come forward, for example, just as leaked documents can shed light on the unknown past.[15] Many exciting testing opportunities therefore exist for historians and social scientists.

Some scholars may nevertheless doubt the value of making and testing predictions about the past. We address three major challenges in turn.[16] First, critics may argue that retrodiction is unhelpful because history is inherently unpredictable and ungeneralizable. In sum, few reliably accurate predictions about the past are possible and, even if some were, the lessons of retrodiction would not translate from one case to another. This kind of approach to history, however, runs into several challenges.[17] We simply do not know if some or all historical subjects are unpredictable or the extent to which some possible insights are generalizable. Some scholars may be able to make more accurate predictions about the past than others, which is exactly what social scientists have found in their studies of competitive forecasting.[18] Making and testing predictions about the past would therefore allow us to determine whether retrodiction is viable and whether it can be generalizable. Furthermore, many scholars already make predictions about the past, whether implicitly or explicitly. Bringing these retrodictions to light, understanding how they work, and making serious efforts to test them could help to improve existing historical practice.

Second, some critics may consider more explicit or formal efforts at retrodiction as being of limited value to existing practice. They might argue that scholars have already changed their minds and updated their own work in the light of new evidence.[19] We do not deny that historians and social scientists sometimes revise earlier work but note that such efforts are relatively rare and often implicit, which limits any insights. If we rely only on the finished works of history, we can usually only speculate about how particular interpretations have taken shape. We therefore learn little about what constitutes effective retrodiction and how to improve historical judgement. In addition, critics might claim that when new data does emerge, it already feeds into the work of other historians or social scientists, who can then assess whether earlier predictions

Downloaded from https://academic.oup.com/ahr/article/130/4/1615/8404686 by guest on 12 February 2026

about the past were correct. We accept elements of this claim. Some of the authors of this article have certainly used previously unavailable archival material to challenge or refine earlier interpretations of the past.[20] Nevertheless, existing approaches are often partial and selective in focus. Many flawed predictions about the past remain unaddressed. Moreover, there is typically limited reflection on where or why earlier retrodictions erred. Although it may be convenient to forget about the inaccurate guesses that litter the study of the past, we are much more likely to repeat them if we ignore or downplay these mistakes.

Third, critics of retrodiction may be wary of making mistaken predictions about the past. We understand these concerns but believe them to be largely misplaced. Many scholars already make indirect or implicit inferences about missing data. Producing and testing retrodictions seeks only to make existing historical arguments clearer and more useful. Some predictions about the past may be disproven but this outcome would be nothing new; historians and social scientists regularly challenge the claims of their peers. Identifying mistakes and learning from them is key to improving our historical judgement. We should not criticize scholars for making honest mistakes but recognize that a willingness to make arguments that might turn out to be wrong is a key part of the process by which historical interpretations take shape.

None of this is to deny the philosophical and methodological challenges posed by the testing of historical claims.[21] We accept that not everything can be predicted or tested and appreciate that subjectivity can challenge historical assessment. Nevertheless, the potential difficulties involved should not dissuade scholars from seeking to make and evaluate retrodictive judgements when viable. Similar challenges have failed to stop the growing acceptance of prediction and testing in other disciplines.[22] Much academic support for testing reflects the rewards on offer, not least helping to expose errors and open minds in otherwise polarized debates.[23] Explicit testing can also facilitate "adversarial collaboration," which can improve accountability and credibility, especially in historiographical debates. By encouraging scholars with competing interpretations of the past to agree on the relevance of specific retrodictions and the terms for testing, future archival discoveries can be used more effectively to encourage scholarly consensus.[24] The insights generated by retrodictive testing on mutually agreed terms could therefore be profound.

The collection and testing of experts' claims about the future have been a crucial component of understanding and improving the art and science of prediction.[25] If we seek to understand whether accurate predictions about the past are possible, and whether some scholars or strategies might fare better than others, historians and social scientists must also find methods to make and test retrodictions in ways that

Downloaded from https://academic.oup.com/ahr/article/130/4/1615/8404686 by guest on 12 February 2026

enable the collection and interpretation of data. We introduce four main approaches to retrodiction for scholars to consider.

First, historians and social scientists could produce a paragraph towards the end of an article or create an online appendix, for example, providing specific and testable predictions about the past, offering levels of confidence in these predictions and explaining how they were reached. Future scholarship could then return to these retrodictions to learn more about the power and validity of the underlying historical argument. The political scientists Randall Schweller and William C. Wohlforth's research concerning the drivers of Soviet retrenchment at the end of the Cold War, for example, provides a series of specific predictions in advance of the release of more archival data, allowing them and other scholars to assess their broader interpretation of events as more evidence emerges.[26] Although this kind of approach is limited in that it does not test any retrodictions, it helps scholars by encouraging them to think carefully about their assumptions and supports scholarship more broadly by facilitating future testing. This approach also shows that making explicit and testable retrodictions need not represent a radical shift or arduous addition to existing research practices.[27]

Second, scholars could prepare and publish a pre-registration report. Historians or socials scientists seeking to test specific hypotheses or advancing broader interpretations of history would begin their research projects as normal, drawing on available evidence and their own expertise, but pause before conducting key archival or interview research. At this point, they would produce a report that outlined their predictions about the past alongside their confidence levels and rationales. Once this kind of report was published, historians and social scientists would venture to the archives or conduct interviews to test their claims. This approach would help other scholars assess the quality of a particular historian's arguments by seeing how well the retrodictions generated by those arguments stand up in the light of evidence that subsequently emerges. Pre-registration reports have become increasingly important in the physical and social sciences by promoting deliberate and considered thinking about the unknown while also limiting many of the problems associated with post-hoc explanation.[28]

Third, historians and social scientists can proactively identify retrodictions made by other scholars in published work and then assess their historical judgement when missing data eventually comes to light. This approach is widely accessible; anyone with access to works of history and the ability to locate new evidence can perform retrodictive testing. Library shelves and online repositories are alive with testing opportunities. Such an approach, however, faces several limitations. Even if retrodictions can be identified and tested they may be poorly specified, making them difficult to evaluate and easy to defend with post-hoc justifications. Without further context, it may also be challenging to discern historians' rationale for, or confidence in, their predictions about the past. Moreover, and relevant also to the previous approaches

Downloaded from https://academic.oup.com/ahr/article/130/4/1615/8404686 by guest on 12 February 2026

introduced above, testing a limited number of retrodictions may help us to assess the work of a particular historian and their specific claims but broader generalizations about a subject or method will naturally be limited.

Fourth, historians and social scientists can devise retrodiction tournaments. Much of what social scientists have learnt about prediction derives from careful evaluation of forecasting tournaments held in the 1980s and 1990s. In these events, invited experts attempted to predict the outcomes of a wide range of global events—from interstate violence to leadership changes—generating nearly 30,000 forecasts, which were scored by their accuracy.[29] Scholars have continued to build on these efforts in the following decades, learning more about the successes and failures of expert predictions and their underlying rationales.[30]

Retrodiction tournaments would involve inviting historians and social scientists to make predictions about data from the past which are unknown to them but are known to the organizers of the event. One way of managing such an event would be for the organizers to collect a series of recently released archival documents, for example, and then provide participants with contextual information pertaining to those documents. Participants would then be invited to use their historical expertise to predict what the documents would show. Organizers could provide historians and social scientists with a small number of choices to help guide their retrodictions or leave the process more open-ended depending on what they wanted to test. The organizers would record participant's retrodictions—alongside their explanations, assumptions, methods, rationale, and confidence levels—to assess their performance and learn more about what make some predictions about the past more successful than others.

Retrodiction tournaments could run synchronously or asynchronously and take place either online or in-person, perhaps as special sessions within annual conferences or larger multi-day workshops. These events could be private or public facing.[31] The latter might include leaderboards tracking claims by different historical scholars. In contrast to other forms of making and testing retrodictions, tournaments are more demanding in terms of organization and rely on securing participants. Nevertheless, they would allow organizers to manage and specify the testing process across larger sample sizes. Retrodiction tournaments would therefore make evaluation easier and generalizations more plausible. In addition, they can facilitate a wide variety of participants, including different types of historians, social scientists, and non-experts, expanding and enriching our understanding of efforts to predict the past.

In all four approaches, we would encourage historians and social scientists to think carefully and explicitly about the confidence levels associated with any retrodictions, whether in terms of their own claims or those of others. The simplest predictions state a single expected outcome, with success measured in binaries: right or wrong, with many

Downloaded from https://academic.oup.com/ahr/article/130/4/1615/8404686 by guest on 12 February 2026

ways to be wrong, and only one to be right. Predictions that come with a stated degree of confidence can be more useful than predictions alone because they give valuable information about the strength or precision of a specific explanation or theory.[32]

Making and testing predictions remains relatively rare in the discipline of history but is more common in fields like psychology and political science.[33] The lessons learned from decades of research in other fields, especially the social and behavioral sciences, can usefully be applied by historians and historically minded scholars to better understand and refine how they make and test their own predictions about the past. Scholars have found that predicting the future is difficult; studies have repeatedly shown that most experts' predictions are not particularly accurate.[34] A series of recent forecasting initiatives assessed the accuracy of social scientists' predictions regarding how societal levels of phenomena like depression, prejudice, and political polarization would unfold in the first year of the COVID-19 pandemic. It found that expert forecasts were no more accurate than those of a sample of average Americans.[35]

We should not therefore assume high levels of retrodictive accuracy in the work of historians or social scientists. Nevertheless, past studies of expert prediction point to practices that can enhance the accuracy of forecasting. The *Good Judgment Project* discovered that a subset of lay forecasters, termed "superforecasters," consistently outperformed experts through methodical analysis, probabilistic thinking, and readiness to revise predictions as new information became available.[36] In sum, the virtue of intellectual humility helped to improve accuracy.[37] In a similar vein, recent studies have found that teams making more accurate forecasts had relevant expertise, were interdisciplinary, used simpler models, and based predictions on prior data.[38] Whether adopting some of these approaches or traits would help historians or social scientists to improve the accuracy of their retrodictions remains an exciting area of future research.

Research from the social and behavioral sciences also highlights the importance of recognizing our own biases in retrospective judgments, which provides an opportunity to adopt strategies that can improve historical expertise. One such strategy is "dialectical bootstrapping," which would entail making an initial prediction about the past and then critically reflecting on why it may be wrong, whether by adopting competing assumptions or using different data, thereby encouraging historians and social scientists to probe biases in their initial approach.[39] Stepping back and trying to make historical judgments from a different perspective can increase objectivity about one's assumptions and raise awareness of other interpretations, thereby leading to more accurate judgments.[40] These kinds of strategies benefit from teamwork. Whereas historians often work alone or in small

Downloaded from https://academic.oup.com/ahr/article/130/4/1615/8404686 by guest on 12 February 2026

groups, research suggests there can be wisdom in crowds. Scholars have repeatedly found that aggregating many individuals' independent estimates and then taking the central tendency typically yields more accurate answers than any individual produces.[41] This result may occur because individuals often possess unique knowledge, such that when enough estimates are averaged, errors cancel each other out, but true, perceptible signals remain.

One of the central insights from decades of research and tournaments testing the accuracy of predictions is that *what* experts think matters far less than *how* they think.[42] To illustrate this point, social scientists often turn to Isaiah Berlin's famous distinction between foxes and hedgehogs.[43] Foxes represent people more interested in using different ideas and approaches to explain events depending on what they think works best. Hedgehogs are people who use one big idea or overarching ideology to understand and explain the world. Although it might be more helpful to treat these ideas as existing on a spectrum rather than simply a dichotomy, the distinction serves a useful purpose. These two types of experts were statistically distinguishable; hedgehogs failed to do better than random guessing while foxes displayed some modest foresight.[44]

Social scientists have shown that this distinction has repeatedly proven more important than professional background or political beliefs to predictive accuracy.[45] Whether foxes or hedgehogs are better at predicting the past remains unclear and therefore represents a fascinating line of future research for scholars.[46] Will diffident historians and social scientists, those that recognize complexity and are willing to adjust their beliefs and ideas, outperform their more confident peers, those that prefer sweeping interpretations or grand theories of the past?[47]

Broad interpretations or grand theories can represent an attractive proposition: by understanding key elements of the world, they can logically and consistently fill many evidentiary gaps in the historical record. Many scholars studying the past have produced work guided by grand theories. The historian Eric Hobsbawm, for example, was an avowed Marxist.[48] Hobsbawm argued that his worldview helped him to understand the past. Critics have instead argued that Hobsbawm's Marxist beliefs clouded his historical judgement.[49]

A valuable way to resolve debates about Hobsbawm's historical accuracy—and thus learn more about the judgement of prominent hedgehogs—would be to identify and assess retrodictions in his work, especially those on which key arguments may rest. As one political scientist hedgehog has put it, "to assess any theory" one can ask, "how well does it explain the past?"[50] We suggest that a good theory or interpretation should be able to explain both the known and unknown past, the latter allowing us to limit many of the problems associated with post-hoc explanation. In addition to assessing the accuracy of scholars' key works, we should also invite larger numbers of historians

Downloaded from https://academic.oup.com/ahr/article/130/4/1615/8404686 by guest on 12 February 2026

and social scientists from similar traditions to participate in retrodiction tournaments, which would allow us to cautiously generalize about the retrodictive power of their worldviews in a more controlled environment.

The accumulation of decades of previously unavailable evidence should allow us to perform tests on many existing predictions about the past. Hobsbawm touched on a similar line of thinking when reflecting on the history of the Russian Revolution after the opening of the Soviet archives in the 1990s. He noted that "during practically all the life of the USSR much was inaccessible, hidden behind locked archive doors and barricades of official lies and half-truths" but "much of what actually happened can now be known because information is available."[51] Hobsbawm appeared pessimistic about historians' accuracy concerning Russian history and believed that "an enormous mass of the literature that appeared during that time will now have to be junked, whatever its ingenuity in using fragmentary sources and the plausibility of its guesswork … When better or more complete data are available, they must take the place of poor and incomplete ones."[52]

Hobsbawm's pessimism about efforts to predict the past is consistent with social scientists' findings concerning the inaccuracy of expert's predictions of the future. It is reasonable to believe that many predictions about the past will likely be proven incorrect when they are eventually tested against better or more complete data. Nevertheless, identifying and testing retrodictions, especially in large numbers, presents a valuable opportunity. Rather than junking old research and selectively replacing existing work, we can instead examine this data to learn from scholars' predictions about the past. By studying many examples of retrodiction, whether successful or not, we can begin to learn more about whether certain approaches or methods can improve historical judgement. Have foxes been more successful in predicting the past than hedgehogs, for example, or have certain kinds of methodological expertise proven more effective than others? Over time, we may be able to identify patterns of effective prediction of the past and, in turn, offer useful insights for future researchers that can help to improve the practice of retrodiction.

Retrodiction presents many valuable opportunities for historians and social scientists. Making explicit claims about the unknown past and then testing them can help us to improve historical judgement by better understanding the extent to which prediction is possible and whether some methods or strategies fare better than others. Retrodiction can also inform scholarly practice. Just as we expect scholars who study the past to reflect on the collection and interpretation of evidence, we should also expect them to think carefully and write clearly about how they will handle gaps in the historical record. Thoughtful and explicit claims about the unknown past can help to raise confidence in scholars'

Downloaded from https://academic.oup.com/ahr/article/130/4/1615/8404686 by guest on 12 February 2026

methods and arguments by promoting clarity and accountability in their work.

Retrodictions can empower the work of historians and social scientists individually but also collectively. The more detailed predictions about the past that we can collect and test, the more we can learn about the art and science of retrodiction. We have introduced and outlined different ways to make and test predictions about the past. Scholars can choose to publish their retrodictions and invite others to test them or produce pre-publication reports and conduct the research themselves. Alternatively, they can identify retrodictions in published work and use new evidence to assess their accuracy. Some scholars may instead prefer larger-scale testing, deciding to organize retrodiction tournaments involving many participants. Collecting and interpreting data from all these approaches can help us to improve our understanding of retrodiction. The longer-term stakes of this project are significant. Scholars in many other fields have made great strides in understanding and improving predictions about the future. These efforts have in turn promoted good practice and helped to improve scholarly judgement. Making and testing retrodictions will allow historians to improve their own historical judgement and our collective understanding of the past.

David James Gill is professor of International Relations at the University of Nottingham, UK. He has published research in the fields of history, international relations, and management with articles appearing in the *Economic History Review*, *Security Studies*, and *Strategic Management Journal*. David has written two books, *Britain and the Bomb* (Stanford University Press, 2014) and *The Long Shadow of Default* (Yale University Press, 2022), and has co-authored one book, *Divided Allies* (Cornell University Press, 2019).

Marc Trachtenberg, an historian by training, is a professor emeritus of political science at UCLA. He is the author of a number of books and about a hundred articles, dealing mostly with twentieth century international politics. His book on method, *The Craft of International History,* was published by Princeton University Press in 2006.

Michael J. Gill is an associate professor of Organisation Studies at the University of Oxford, UK. His research interests lie in using historical and qualitative research methods to understand individuals in the workplace. He has published in leading journals in the field of business and management including the *Academy of Management Annals, Academy of Management Review, Journal of Management*, and *Strategic Management Journal*.

Philip E. Tetlock holds the Penn Integrates Knowledge (Annenberg) chair at the University of Pennsylvania, with cross-appointments at Wharton and the School of Arts and Science. His work crosses the

Downloaded from https://academic.oup.com/ahr/article/130/4/1615/8404686 by guest on 12 February 2026

boundaries of cognitive, political and organizational science and addresses topics such as accountability, value conflict, counterfactual reasoning and taboo trade-offs. His current work focuses on forecasting tournaments and their potential to improve intelligence analysis and to depolarize unnecessarily polarized debates on domestic as well as national security issues.

Thomas K. Robb is a senior lecturer in History at Oxford Brookes University, UK. He has written and co-authored several articles and books, with the most recent being *Divided Allies* (Cornell University Press, 2019). He researches on areas of US strategic history and is writing two books. The first focuses on the CIA and presidential decision-making and the second focuses on the interplay between economics and military decision-making during the US Civil War.

Michael E. W. Varnum is professor of Psychology at Arizona State University. His work has appeared in leading scientific journals including *Nature Human Behaviour*, *Proceedings of the National Academy of Sciences*, *PNAS Nexus*, and *Proceedings of the Royal Society B: Biological Sciences*.

Cendri A. Hutcherson is an associate professor of Psychology at the University of Toronto Scarborough, Canada. She takes a computational approach to decision making, information sampling, and scientific predictions, with work appearing in some of the leading journals in the fields of psychology and neuroscience, including *Nature Human Behaviour*, *Nature Communications*, *Proceedings of the National Academy of Sciences*, *Neuron*, and *American Psychologist*.

Igor Grossmann is professor of Psychology at the University of Waterloo, Canada, whose interdisciplinary research bridges philosophy, anthropology, economics, and psychology to study human judgment and wisdom. He developed the Common Wisdom Model and leads the international Wise Judgment Consortium to challenge Western-centric views of decision-making. An elected Member of the College of the Royal Society of Canada and recipient of major awards from the APA and SPSP, Grossmann also founded the Forecasting Collaborative, co-hosts the *On Wisdom Podcast*, and leads initiatives such as World-after-COVID and Futurescape.

Zoe Trodd is professor of Modern Slavery and Human Trafficking, and director of the Rights Lab, at the University of Nottingham, UK. Her books include *Picturing Frederick Douglass: An Illustrated Biography of the Nineteenth Century's Most Photographed American* (Norton, 2015), *The Tribunal: Responses to John Brown and the Harpers Ferry Raid* (Harvard University Press, 2012), and *To Plead Our Own Cause: Personal Stories by Today's Slaves* (Cornell University Press, 2008).

Downloaded from https://academic.oup.com/ahr/article/130/4/1615/8404686 by guest on 12 February 2026

1 Scholars have used different words or phrases to categorize prediction of the past rather than the future. On retrodiction as a form of "prediction" that "comprises statements about the past" see Karl Popper, *The Logic of Scientific Discovery* (Routledge, 1992 [1959]): 38n2. On retrodiction as a method used to "fill in the gaps," and as a method often used "unconsciously," see Paul Veyne, *Writing History: Essay on Epistemology* (Wesleyan University Press, 1984): 144.

2 See, for example, Philip E. Tetlock, *Expert Political Judgement* (Princeton University Press, 2017); Philip E. Tetlock and Dan Gardner, *Superforecasting: The Art and Science of Prediction* (Crown Publishers, 2015).

3 Our focus in this article is largely on qualitative historical research. For a quantitative approach to making predictions about long-run economic growth in history, for example, see Meredith M. Paker, Judy Z. Stephenson, and Patrick Wallis, (2025) "Predictive Modeling the Past," Economic History Working Papers 379 (2025).

4 Scholars have made similar claims about counterfactual reasoning. As Martin Bunzl explains, counterfactuals "are not as easy to avoid in the practice of history as one might think" and "plays an unavoidable implicit role in history." See Martin Bunzl, "Counterfactual History: A User's Guide," *American Historical Review* 109, no. 3 (2004): 845, 857. Nevertheless, there are important differences. Whereas counterfactual reasoning seeks to create and understand hypothetical histories, retrodiction seeks to accurately predict the factual past. See also Philip E. Tetlock and Richard Ned Lebow, "Poking Counterfactual Holes in Covering Laws: Cognitive Styles and Historical Reasoning," *American Political Science Review* 95, no. 4 (2001): 829–43.

5 Robert Self, *Britain, America, and the War Debt Controversy* (Routledge, 2006), 214.

6 David James Gill, *The Long Shadow of Default: Britain's Unpaid War Debts to the United States, 1917-2020* (Yale University Press, 2022).

7 Critiques of the explanatory approach in the behavioral sciences suggest considerable danger lurks in overreliance on explanation, including a tendency to develop overly complex hypotheses that fail to generalize to new situations, otherwise known as "overfitting." See Robyn M. Dawes, "Prediction of the Future Versus an Understanding of the Past: A Basic Asymmetry," *The American Journal of Psychology* 106, no. 1 (1993): 1–24; Jake M. Hofman, Amit Sharma, and Duncan J. Watts, "Prediction and explanation in social systems," *Science* 355, no. 6324 (2017): 486–88.

8 Carl Hempel, "The Function of General Laws in History," *Journal of Philosophy* 39, no. 2 (1942): 35–48, reprinted in Patrick Gardiner, ed., *Theories of History* (The Free Press, 1959), 344–56; Carl Hempel and Paul Oppenheim, "Studies in the Logic of Explanation," *Philosophy of Science* 15, no. 2 (1948): 138, especially n2a; Marc Trachtenberg, *The Craft of International History: A Guide to Method* (Princeton University Press, 2006), 4. On the relationship between prediction and observation using Einstein's theory of general relativity, for instance, see Imre Lakatos, "Lectures on Scientific Method," in Imre Lakatos and Paul Feyerabend, *For and Against Method: Including Lakatos's Lectures on Scientific Method and the Lakatos-Feyerabend Correspondence*, ed. Matteo Motterlini (University of Chicago Press, 1999), 99–100.

9 Trachtenberg, *Craft of International History*, 4.

10 For a summary and synthesis of competing explanations of the Soviet Union's demise supported by extensive research, see Vladislav Zubok, *Collapse: The Fall of the USSR* (Yale University Press, 2022). An explicit example of scholars studying Soviet decision making and advancing retrodictions about what future archival data should reveal can be seen in Randall L. Schweller, and William C. Wohlforth, "Power test: Evaluating Realism in Response to the End of the Cold War," *Security Studies* 9, no. 3 (2000), 60-107.

11 John R. Platt, "Strong Inference: Certain Systematic Methods of Scientific Thinking May Produce Much More Rapid Progress Than Others," *Science* 146, no. 3642 (1964): 347–53.

12 Neal J. Roese and Kathleen D. Vohs, "Hindsight Bias," *Perspectives on Psychological Science* 7, no. 5 (2012): 411–26. As Aroop Mukharji and Richard Zeckhauser explain, "once events have occurred, people armed with the knowledge of hindsight tend to exaggerate those events' predictability and inevitability." Mukharji and Zeckhauser continue by suggesting "that *all* individuals, including professional historians, underplay uncertainty to some degree when analyzing historical developments." See "Bound to Happen: Explanation Bias in Historical Analysis," *Journal of Applied History* 1, no. 1-2, 5-6.

13 Welton Chang, Pavel Atanasov, Shefali Patil, Barbara A. Mellers, and Philip E. Tetlock, "Accountability and Adaptive Performance Under Uncertainty: A Long-Term View," *Judgment and Decision Making* 12, no. 6 (2017): 610–26.

14 See, for example, Sheila Fitzpatrick, "Impact of the Opening of Soviet Archives on Western Scholarship on Soviet Social History," *The Russian Review* 74, no. 3 (2015): 377–400.

15 See, for instance, https://www.archives.gov/research/pentagon-papers.

16 For a response to criticisms of prediction and objectivity in general, see Tetlock, *Expert Political Judgement*, 216–38.

Downloaded from https://academic.oup.com/ahr/article/130/4/1615/8404686 by guest on 12 February 2026

17 On the problems of nomothetic-idiographic distinctions, see Paul W. Schroeder, "History and International Relations Theory: Not Use or Abuse, but Fit or Misfit," *International Security* 22, no. 1 (1997): 66.

18 Tetlock and Gardner, *Superforecasting*.

19 Veronica Wedgwood rewrote her book on the life of the Earl of Strafford, for example, out of a respect for the evidence that emerged over thirty years. See Richard Cohen, *Making History: The Story Tellers Who Shaped the Past* (Simon & Schuster, 2022). For the texts, see C. V. Wedgwood, *Strafford, 1593–1641* (Jonathan Cape, 1935); *Thomas Wentworth, First Earl of Strafford, 1593–1641: A Revaluation* (Jonathan Cape, 1961). As another example, John Lewis Gaddis updated some of his beliefs concerning the Cold War based on the emergence of new evidence. See *We Now Know: Rethinking Cold War History* (Oxford University Press, 1997).

20 See, for example, Marc Trachtenberg, *A Constructed Peace: The Making of the European Settlement, 1945–1963* (Princeton University Press, 1999); Thomas K. Robb and David James Gill, *Divided Allies: Strategic Cooperation against the Communist Threat in the Asia-Pacific during the Early Cold War* (Cornell University Press, 2019).

21 On the challenges and rewards of "quantifying the unquantifiable," see Tetlock, *Expert Political Judgement*, esp. 1–19, 216–38. For historians' complex and evolving understanding of objectivity, for example, see Peter Novick, *That Noble Dream: The 'Objectivity Question' and the American Historical Profession* (Cambridge University Press, 1988).

22 As one of the authors of this article has previously written, "The arc of history has already begun bending toward tournaments." See Tetlock, *Expert Political Judgement*, xxviii.

23 Dan Gardner, *Future Babble: Why Expert Predictions Fail and Why We Believe Them Anyway* (Virgin Books, 2011); Tetlock, Mellers, Rohrbaugh, and Chen, "Forecasting Tournaments."

24 Cory J. Clark and Philip E. Tetlock, "Adversarial collaboration: The next science reform." In Craig L. Frisby, Richard E. Redding, William T. O'Donohue, & Scott O. Lilienfeld (eds.), *Ideological and Political Bias in Psychology: Nature, Scope, and Solutions* (Springer, 2023).

25 For classic texts, see Tetlock, *Expert Political Judgement*; Tetlock and Gardner, *Superforecasting*; Gardner, *Future Babble*; Nate Silver, *The Signal and the Noise: The Art and Science of Prediction* (Allen Lane, 2012).

26 Schweller and Wohlforth, "Power Test," especially 102–106.

27 For another and more recent example, which focuses on why the United Kingdom repaid its post-war debts to the United States, see David James Gill, "War Debts and the Repayment Norm," *Security Studies*, Vol. 34, No 3 (2025): 489-90.

28 Brian Nosek, Charles Ebersole, Alexander DeHaven, and David Mellor, "The Preregistration Revolution," *Proceedings of the National Academy of Sciences* 115, no. 11 (2018).

29 Tetlock, *Expert Political Judgement*, xix.

30 There exists a growing body of literature concerning the assessment of expert performance. See, for example, Barbara A. Mellers, Lyle Ungar, Jonathan Baron, Jamie Ramos, Burcu Gurcay, Katrina Fincher, Sydney Scott, Don Moore, Pavel Atanasov, Samuel Swift, Terry Murray, Eric Stone, and Philip E. Tetlock, "Psychological Strategies for Winning a Geopolitical Forecasting Tournament," *Psychological Science* 25, no. 5 (2014): 1106–115; Philip E. Tetlock, Barbara A. Mellers, Nick Rohrbaugh, and Eva Chen, "Forecasting Tournaments: Tools for Increasing Transparency and Improving the Quality of Debate," *Current Directions in Psychological Science* 23, no. 4 (2014): 290–95; Philip E. Tetlock, Christopher Karvetski, Ville A. Satopää, and Kevin Chen, "Long-range subjective-probability forecasts of slow-motion variables in world politics: Exploring limits on expert judgment," *Futures and Foresight Science* 6, no. 1 (2014): e157; Igor Grossmann, Michael Varnum, Cendri Hutcherson, and David Mandel, "When Expert Predictions Fail," *Trends in Cognitive Sciences* 28, no. 2 (2024): 113–23.

31 For more on the forecasting tournaments that have been run to predict future geopolitical events, see Tetlock et al., "Forecasting Tournaments," 290–95.

32 On confidence and accuracy, see Tetlock, and Gardner, *Superforecasting*, 138-39, 181. An appreciation of confidence levels allows historians to assess the comparative strength of evidence for different explanations. Although retrodictions are not frequently generated by statistical or algorithmic methods and therefore do not typically come with explicit probability estimates, new methods and approaches suggest it could become feasible. See Joseph Risi, Amit Sharma, Rohan Shah, Matthew Connelly, and Duncan J. Watts, "Predicting History," *Nature Human Behaviour* 3, no. 9 (2019): 906–12.

33 Igor Grossmann, Michael E. W. Varnum, Cendri A. Hutcherson, and David R. Mandel, "When Expert Predictions Fail," *Trends in Cognitive Sciences* 28, no. 2 (2024): 113–23; Tetlock and Gardner, *Superforecasting*.

34 Tetlock, *Expert Political Judgment*.

35 The Forecasting Collaborative, "Insights into the Accuracy of Social Scientists' Forecasts of Societal Change," *Nature Human*

Downloaded from https://academic.oup.com/ahr/article/130/4/1615/8404686 by guest on 12 February 2026

Behaviour 7, no. 4 (2023): 484–501. See also Cendri A. Hutcherson, Konstantyn Sharpinskyi, Michael EW Varnum, Amanda Rotella, Alexandra S. Wormley, Louis Tay, and Igor Grossmann, "On the Accuracy, Media Representation, and Public Perception of Psychological Scientists' Judgments of Societal Change," *American Psychologist* 78 (2023): 968–81.

36 Tetlock, and Gardner, *Superforecasting*.

37 Tenelle Porter, Abdo Elnakouri, Ethan A. Meyers, Takuya Shibayama, Eranda Jayawickreme, and Igor Grossmann, "Predictors and Consequences of Intellectual Humility," *Nature Reviews Psychology* 1, no. 9 (2022): 524–36.

38 Forecasting Collaborative, "Insights into the Accuracy of Social Scientists' Forecasts of Societal change." See also Hutcherson et al., "On the Accuracy, Media Representation, and Public Perception of Psychological Scientists' Judgments of Societal Change," 968–81.

39 Stefan M. Herzog and Ralph Hertwig, "Think Twice and Then: Combining or Choosing in Dialectical Bootstrapping?" *Journal of Experimental Psychology: Learning, Memory, and Cognition* 40, no. 1 (2014): 218–32.

40 Igor Grossmann, Anna Dorfman, Harrison Oakes, Henri C. Santos, Kathleen D. Vohs, and Abigail A. Scholer, "Training for Wisdom: The Distanced-Self-Reflection Diary Method," *Psychological Science* 32, no. 3 (2021): 381–94.

41 James Surowiecki, *The Wisdom of Crowds* (Anchor Books, 2005).

42 Tetlock, *Expert Political Judgement*, 2. Decades of testing and reflection have generated many other valuable insights about the kinds of strategies or approaches scholars can use to improve the accuracy of their predictions about the future. For an overview of key guidelines, see Tetlock and Gardner, *Superforecasting*, 277–86. As the authors warn, however, we should be cautious of any commandments or binding rules: "Guidelines are the best we can do in a world where nothing is certain or repeatable." Ibid., 286.

43 Isaiah Berlin, "The Hedgehog and the Fox," in *The Proper Study of Mankind: An Anthology of Essays* (Chatto & Windus, 1997), 436–498.

44 Tetlock and Gardner, *Superforecasting*, 67-68; Gardner, *Future Babble*, 27-28.

45 Berlin, "The Hedgehog and the Fox," in *The Proper Study of Mankind*, 436–98.

46 Some historians such as Donald Kagan have already advanced an answer: "Historians should, in the first instance, be foxes … [but] it is this mixed path taken by the historian, chiefly of the fox but with a necessary element of the hedgehog, that I believe promises the best results." *On the Origins of War and the Preservation of Peace* (Doubleday, 1995), 9.

47 Tetlock, *Expert Political Judgement*.

48 For an example of his work, see Eric Hobsbawm, *The Age of Extremes: The Short Twentieth Century, 1914–1991* (Viking Penguin, 1994). Historians are not unique in the use of grand theories to study the past. The political scientist John J. Mearsheimer, for example, is a self-described neorealist. Like Hobsbawm, he has claimed that his worldview has helped him to understand the past while critics have argued that neorealist theories have clouded historical judgement. See John J. Mearsheimer, *The Tragedy of Great Power Politics* (W.W. Norton, 2001). For criticisms of Mearsheimer's approach, see Jonathan Haslam, "John Mearsheimer's 'Elementary Geometry of Power': Euclidean Moment or an Intellectual Blind Alley?" in Ernest R. May, Richard Rosecrance, and Zara Steiner, eds., *History and Neorealism* (Cambridge University Press, 1999). On Mearsheimer's predictive track record, see Philip E. Tetlock, "Tracking Forecasting Accuracy of Geopolitical Schools of Thought—and Causes of their Predictive Successes and Failures," *Critical Review* 36 no. 4 (2024): 515–25.

49 Tony Judt, "The Last Romantic," *The New York Review of Books*, November 20, 2003. Some historians see theories such as Marxism and good historical practice as existing in tension. Richard J. Evans concluded in his biography of Hobsbawm that, "throughout his career as an historian … [he] was pulled one way by his Communist and, more broadly, his Marxist commitment, and another by his respect for the facts, the documentary record and the findings and arguments of other historians." See *Eric Hobsbawm: A Life in History* (Oxford University Press, 2019), 543.

50 John J. Mearsheimer, "Realists as Idealists," *Security Studies* 20 no. 3 (2011): 424.

51 Eric Hobsbawm, *On History* (Abacus, 1997), 242.

52 Hobsbawm, *On History*, 243.

The American Historical Review, 2025, 130(4), 1615–1630
rhaf590
© The Author(s) 2025. Published by Oxford University Press on behalf of the American Historical Association. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Downloaded from https://academic.oup.com/ahr/article/130/4/1615/8404686 by guest on 12 February 2026