Counterfactualism and Anticipation


Neville Morley, Professor of Classics & Ancient History, University of Exeter, UK

Abstract

Most attempts at forecasting the future depend, explicitly or implicitly, on knowledge about the past, whether this is then used to offer possible analogies or to support normative theories with data about past events and trends. This approach is open to criticism both on the grounds of its assumptions about continuity and a tendency towards deterministic thinking, and on the grounds that our knowledge of the past is less secure and more discursive than such attempts at prediction assume. Counterfactualism, the development and exploration of accounts of ‘what might have been’ – which can be focused on obtaining better understanding of the past, or on refining theories of the present, or on speculations about the future – offers an alternative approach that emphasises the openness of historical developments. Its primary role is not to improve forecasting but to highlight its limitations, to expand our knowledge of how humans think about the future and the cognitive biases that dominate such thinking, and to establish the ethical imperative of engaging with possible futures. The qualities which make counterfactualism a marginal and suspect activity within historiography and social science are precisely those which make it an essential aspect of the discipline of anticipation.

Introduction

Many attempts at forecasting or predicting the future rely on knowledge about the past. This knowledge may be in the form of detailed and reliable data from the recent past, interpreted through existing theories and models, as in the cases of weather forecasting and climate science. The main issue for this type of forecasting is the non-linearity of the processes it seeks to predict (the chaotic nature of weather systems, and many other natural phenomena) and hence the uncertainty of any such forecasts, especially predictions of specific developments rather than general trends, beyond the short term.

Such approaches are seen to be less effective in the case of more ‘open’ systems, where processes and developments may be shaped by any number of different factors, and hence appear to be even less predictable than natural systems; this applies above all to different aspects of human society. One response to this problem has been a focus on different forms of abstract modelling, for example in various fields of economics, to explore the interaction of a limited number of key variables, ‘all other things being equal’; this constitutes a powerful intellectual tool, but one which generates knowledge that at best has only a partial connection to the ‘real’ social world, open, and under-determined, and whose perceived relevance to the future rests on an assumption of continuity in all variables other than those studied in the model (Elster 2015).

An alternative approach has therefore been to seek to develop a richer idea of the dynamics of social behaviour, especially for those areas of human life that do not seem to lend themselves to data-driven analysis or abstract modelling (this discussion will take as a key example a theme in contemporary global politics), by looking in more detail at historical events, on the assumption that ‘what happened’ is a solid basis for forecasting what is likely to happen in future (Allison & Ferguson 2016).
There are, however, significant problems with this form of historical argument, and above all the assumption that our knowledge of the past can offer such an objective foundation for evaluation of present and future prospects. It therefore remains a minority approach within mainstream social science. Above all, attempts at ‘learning from the past’ face the problem that detailed historical research tends more and more to emphasise the non-comparability of different events, processes and societies; this means that it is necessary for such comparative/analogue analysis to smooth over or occlude the effects of change over time in order to make past events into a usable resource for analysis of the present and future, even at the expense of reducing the credibility of claims about the specific, objective and ‘real’ nature of the historical data that supposedly form the foundation of such arguments.

At the same time, research in the theory and philosophy of history reveals how far all historical events are open to redescription according to present assumptions and agendas, rather than having an objective existence as potential analogues for present and future developments. Finally, the majority of attempts at identifying patterns in the past that can potentially be extended into the future assume linear models of historical change. In other words, the majority of such attempts at predicting future developments on the basis of historical information are recuperated back into the law-based social science model, but with the relevant laws and principles left largely implicit and taken-for-granted rather than properly specified or explicitly analysed.

But this is not to say that contemplation of the past has no value for considerations of possible futures; we simply need to view the past differently, in a manner which recognises the problems and limited utility of conventional approaches, and we need to modify our expectations of what we can learn from the past. One aspect of this is a better understanding of how humans think about past, present and future, and the nature and effects of the different stories we tell about the connections between them (cf. Liveley 2017). Another is the development of a different way of analysing past events, in order to develop a sense of the openness and instability of historical developments; and this is the realm of counterfactualism.

The Problem of the Past

A Russian joke from the Soviet era, following a standard format, runs as follows:

Radio Armenia was asked: Is it possible to foretell the future? Radio Armenia answered:
Yes, no problem. We know exactly what the future will be. Our problem is with the past; that keeps changing. (Judd 2010: 830)

The point is obvious, a mockery of the claims of official Marxist-Leninist theory to possess certain knowledge of the dynamics of historical processes and the inexorability of progress, and a satire on the instability of the past as it gets rewritten to conform to present circumstances. It reverses the expected order of things, in which the past is a known, fixed object because it has already happened, while the future is unknown. In its scepticism about official narratives, however, the joke raises the spectre of unavoidable uncertainty about past, present and future; all we really know is that these are malleable, and that what can be said of them (now, or in future) is subject to the whims of political authority, even if those contradict the narratives of our own lived experience. The only safe assumption on which to anticipate the future is that anything might happen, even if that cannot be openly admitted.
Within the contemporary western tradition of social science, there has at times been a similar confidence – though not officially endorsed or imposed in the same manner – in the capacity of nomothetic approaches (that is, those focused on general scientific laws) to predict future developments. However, this confidence is grounded in a belief in the fixed and knowable nature of the past and the present, established through empirical study of historical evidence and current data rather than determined or distorted by present-day ideology, which can then serve as a foundation for the development of these explanatory and predictive theories (Pearl 2000).

Belief in the possibility of learning from the past is of course far older; it dates back to the fifth-century BCE Greek historian Thucydides, who declared that his account of a war between Athens and Sparta (known as the Peloponnesian War) would be a “possession for ever”, and that he would be content “if it is found useful by those who want exact knowledge of what happened in the past and what, the human condition being what it is, will happen again in the same or similar manner” (Thucydides 1.22.3; see Morley 2014). This idea formed the basis for a centuries-long tradition of ‘exemplary’ history, in which past individuals, actions and events were held up as models to be studied and imitated in the present.

From the nineteenth century onwards, a sense developed among scholars of the vast differences between modern society and economy and those of previous eras, and an expectation that the future would be more different still; a growing gap between the ‘space of experience’ and the ‘horizon of expectation’, in the terms coined by the intellectual historian Reinhart Koselleck (1985). This clearly cast doubt on the project of seeking direct analogies between past and present as a basis for understanding and prediction. However, while it led to the discrediting of traditional exemplary history, this new perspective did not render the past irrelevant; it simply directed historians, and still more social scientists, towards the task of identifying the underlying continuities (for example, economic laws or human psychology) that meant historical instances were still considered comparable, if no longer directly analogous. Indeed, Thucydides was established as a founding figure in this enterprise, with his claim that “the human condition” ensured a degree of repetitiveness in historical development; he could be represented as a pioneering social scientist rather than a historian in any conventional sense.

In any case, appeal to past precedent continues to carry significant rhetorical power, whether or not it is advanced as part of a nomothetic approach. It allows the framing of simple conditionals to indicate or advocate a course of future action: on the basis of this past event, if we do X then the result will be Y, or if we fail to do X then Z will happen. If we embark on this programme of fiscal stimulus, we will lift the economy out of depression. If we do not act to stop this dictator, he will become bolder and bring war to the entire region.

These two examples are not identical, and not only because one is framed in terms of taking a positive step to cause something to happen that would not otherwise have done so while the other counsels action to prevent something from happening that would otherwise have done. The first can call on the authority of Keynesian economic theory, to offer an analytical explanation of past events (the Great Depression), a diagnosis of the present, and a justification for assuming that future events will follow the same pattern. The latter offers only a (familiar and value-laden) narrative of events in the past and the conviction that these will repeat themselves in similar conditions in the future. Both, however, draw rhetorical authority from the fact that what they propose for the future has happened at least once before: the anticipated sequence of events
is known (or believed) to be possible if not in fact inevitable – and indeed the continuing cultural dominance, in Anglophone countries, of particular narratives of the Second World War mean that the untheoretical, even mythical argument carries greater force than the one which appeals to theoretical knowledge as part of its claim to authority. Even if the prediction is framed in less definite terms, to deflect accusations that the future can never be known with such certainty – if we appease this dictator then he may start a war – the fact that this has happened in the past still carries significant weight.

A concrete example of such an attempt at characterising the present situation, forecasting future developments and offering policy advice on the basis of past events is ‘Thucydides’s Trap’, an idea developed by the American international relations theorist Graham Allison (2017). This builds on the argument offered by Thucydides that the truest but least discussed reason that made the Peloponnesian War inevitable was the growth of the power of Athens and the fear this inspired in the Spartans. Building on Thucydides’ claim, mentioned above, that readers will learn from his account true understanding not only of the past but of similar events in future, Allison rephrases this as a normative principle of inter-state relations, that when a ‘ruling power’ is confronted by a ‘rising power’, the likely result is war. Sixteen historical cases studies are adduced, in which a similar dynamic is identified, including the confrontations between France and the Hapsburgs in the sixteenth century, France and Britain from the seventeenth to the early nineteenth centuries and the United States and the Soviet Union after the Second World War, as well as the obvious examples of the First and Second World Wars; in the majority of these cases, the confrontation of ruling and rising power led to war. Finally, Allison offers an interpretation of current relations between the United States and China in similar terms, and hence argues that, on the basis of the historical data and the general principle identified by Thucydides, war must be seen as a likely result unless steps are taken to prevent this.

This policy advice, that the United States should actively seek to calm possible tensions rather than risk unstoppable escalation by complacently assuming that war is now impossible (as some have argued, on the grounds of several decades of globalisation, convergence, and economic interdependence), seems entirely reasonable as a precautionary principle. However, other conclusions might be drawn from this historical data, even if Allison’s historical claims and predictions are accepted at face value – for example, that the likelihood of war makes it imperative to prepare for it, even at the risk of thus making it more likely. More significantly, the evidence of past events is by no means a secure empirical foundation for such an analysis, even if, once again, the indications are that it is rhetorically effective.

It is, as Allison has demonstrated, possible to give accounts of the sixteen different case studies in terms of ‘Thucydides’s Trap’, but that is not the same as saying that these are correct accounts of ‘what really happened’ in the past. Even ‘Thucydides’ account of the causes of the Peloponnesian War can be questioned, on the grounds both of whether this is a correct reading of his actual interpretation and, more importantly, by comparing it with other historical evidence. This is still more true of the other examples, where detailed study strongly suggests that the situations were far more complex and multi-faceted than Allison’s simple characterisation of them suggests. To explain the outbreak of the First World War as the result of rivalry between a ruling United Kingdom and a rising Germany is, while not wholly untrue, certainly a drastic simplification of the complex interaction of different states and alliances, as well as of different
levels of causation (conscious strategies and decision-making, but also longer-term economic, geographical, and technological factors). In brief, the apparently objective data from the past, on the basis of which the present situation is diagnosed as following a familiar and predictable pattern, can always be reinterpreted and redescribed.

Allison’s model is taken here as an example not because it is uniquely flawed, but because its assumptions about the use of historical data are relatively explicit, and so reveal the essential problem in such approaches. The past is assumed to be known, and hence available for use as a basis for analysis and extrapolation; further, historical processes are assumed to be linear, so that we can expect the same results – or at least the same probability of different outcomes – from similar situations. However, more detailed research into the particular circumstances of a given historical event (as one would expect, in order to ground diagnosis and forecasts more solidly in the data) inevitably tends towards the idiographic, a focus on the particular rather than the general. This raises the problem of historicism, whereby each past moment appears ever more unique and incomparable with other events except in the most general terms.

Further, the fact that the consequences of (allegedly) similar past situations can be represented in similar terms is not a sound basis for claims that there is a consistent, linear principle to be discerned in historical events. Comparison of historical case studies of course remains possible, through continued reliance on nomothetic approaches or by making use of Weberian ideal types, but it becomes increasingly difficult to claim that any conclusions, forecasts, or policy recommendations drawn from them are truly falsifiable with reference to historical evidence. At best, we are left with conventional normative theories, and forecasts based on them, that draw rhetorical power but no empirical support from history.

One widespread response to the problem of historicism and the discursive nature of knowledge about the past is to rely solely on general theories and laws. However, it is not clear that such theories, based on the assumption that our understanding of the present is not only reliable and complete but also universally and eternally valid for hypothetical futures, are sufficient for our purposes. They make claims about a predictable future, knowable in advance, that replicate the old idea of a known, singular past even if they do not explicitly reference it. In terms of the tripartite distinction outlined by Poli (2017), they relate to forecasts and foresight, but not to anticipation, which is explicitly non-predictive, qualitative (idiographic) and focused on discontinuity.

What is required is a different approach to understanding the past, and this is where counterfactualism offers possibilities. Counterfactuals are a specific form of conditional statement, a distinctive and familiar part of natural language, even if their truth conditions raise various philosophical problems. In popular terms, they revolve around ‘what if?’ questions; to take the example used by David Lewis, the central figure in philosophical discussions of counterfactualism and possible worlds, “If kangaroos had no tails, they would topple over”. “I believe,” Lewis argued, “and so do you, that things could have been different in countless ways... I therefore believe in the existence of entities that might be called ‘ways things could have been’” (1973: 84). Applied to the study of the past, this approach offers an alternative way of exploring causation and the dynamics of historical processes; it is widely derided and marginalised within mainstream historiography, for focusing on things that did not happen – but that is precisely
why it may be a useful resources for thinking productively about another not (yet) existent thing, the future.

**Past-focused Counterfactuals**

Counterfactuals are often discussed as a single, unified form of analysis and argument, especially but not only when they are explored from within a specific discipline (e.g. Ferguson 1997 and Tucker 1999 and 2016 on historiography; Lebow 2010 on international relations theory. Hawthorn 1991 is striking in his willingness to cross, or simply ignore, the boundary between history and social science). Different typologies can be suggested; for example, the important discussion by Tetlock & Belkin (1996) focuses on the different methods employed in identifying and developing counterfactual arguments, including idiographic, nomothetic and statistical, while acknowledging that the majority of concrete cases combine such approaches, especially by incorporating normative laws into ideographical counterfactuals or by including idiographic descriptions in nomothetic studies.

An alternative approach, which will be adopted here, is to distinguish different counterfactual approaches according to their primary aims; that is to say, what the counterfactual is intended to illuminate. The majority of counterfactuals take the past as their subject matter, exploring how things that have happened could have happened differently; but this may be done in order to improve our understanding of the present and/or the future (that is, counterfactualism as an attempt at recuperating the past for purposes of forecasting), or in order to improve our understanding of the past as an end in itself. The latter is the most familiar form, not least because of its resemblance to the popular genre of ‘alternative histories’ in novels and television series, such as *SS-GB* and *The Man in the High Castle* (or, in a more literary register, Philip Roth’s *The Plot Against America*). The idea that our world could have become quite different if certain key events or decisions had gone differently in the past carries considerable appeal for non-academic audiences – primarily, it would seem, as a form of dystopian fiction in which the ‘wrong’ side won or the ‘wrong’ thing happened.

Academic historians are less overt in their ideological commitments, though the majority of published counterfactual studies have had a significant overlap with fictional attempts (see Tucker 1999), above all in their focus on very traditional forms of narrative history, organised around the results of individual battles and individual decisions, usually in relation to political and military activities. In some cases, at least, this is in part because such accounts of ‘what might have been’ are written for a wider, non-academic audience rather than for other historians. They depend on their readers’ prior knowledge of what did happen, at least in general terms, and hence tend to concentrate on familiar moments in history, understood in a conventional form: what if the Spanish Armada had succeeded, what if Napoleon had lost at Trafalgar, what if Hitler had invaded Britain in 1941? In some cases, however, a turn to counterfactuals and an emphasis on chance and contingency represents a more or less explicit commitment to *l’histoire événementielle* (if not indeed to a version of Thomas Carlyle’s Great Man theory of history): a history focused on short-term events at a human level, reinstating the active role of individuals in history, implying a rejection of forms of historiography that concentrate on structures and longer-term changes, perceived to favour deterministic understanding of historical change (Ferguson 1997 explicitly presents his counterfactual project as anti-Marxist).
The fact that the majority of published counterfactuals have focused on familiar and sometimes trivial themes and presented them in conventional, if not explicitly conservative, forms does not mean that this ideological agenda is intrinsic to the exercise – there are examples in economic and social history, for example on the effects of the Black Death or on the demographic history of Roman Italy (above all Hawthorn 1991, but see also Morley 2001) – but it has undoubtedly contributed to the marginalisation of counterfactual approaches within mainstream historiography. There are other, more substantial arguments against counterfactualism as a historical method (summarised in Evans 2014). History is mainly conceived as the study of the actual, or at least an attempt at getting as close to this as possible on the basis of the evidence of past activity that has survived, whereas counterfactuals explicitly concern themselves with the study of what did not in fact happen (Hawthorn 1991: 10). Logically, things which did not happen cannot have left any evidence of their occurrence, so on what basis can counterfactual accounts be evaluated, other than their qualities as fiction? How does one decide which counterfactuals are worth studying, given that what actually happened is a single unitary phenomenon whereas the list of other things that could have happened is presumably infinite – and, therefore, what assumptions are being smuggled into the argument that some possible alternatives are more plausible or more worth studying than others?

The counter-argument to this critique is that counterfactuals are simply unavoidable in any discussion of historical causation (Sustein 2016). It is simply a question of whether they are discussed explicitly (so that the underlying assumptions about causation and probability can be evaluated) or whether they are left implicit, creating the impression that a specific historical development was necessary or inevitable. “In a non-experimental and non-comparative discipline one can hardly discuss the relative importance of causes without engaging in some kind of thought experiment where one removes successively and separately each of the causes in question and evaluates what difference the absence of this cause would have made to the phenomenon in question.” (Elster 1978: 176). More succinctly, any satisfactory explanation of a historical phenomenon must also account for negatives, that is, why other possible outcomes did not in fact occur (Tucker 2016: 335); counterfactuals are the obvious means for avoiding retrospective assumptions that what did in fact happen was by definition the most likely if not the only possible outcome.

To consider a crude and familiar example, the case made in Allison’s Thucydides’s Trap that the First World War was the result of the rivalry between Britain and Germany implies that (i) such a rivalry could only have led to war, not to other forms of competition; (ii) without the existence of this rivalry, Europe would not have fallen into war; (iii) it was only this rivalry, rather than the conjunction of this rivalry with other continental rivalries, that was crucial. All these assumptions – and many other possibilities – are best explored through posing counterfactual questions. Further, we can observe that such an account, emphasising underlying structural factors and path dependency as the main driver of events as they occurred, is already a kind of counterfactual, developed in opposition to the conventional narrative account of the outbreak of war: the assassination of Archduke Ferdinand in Sarajevo as a trigger for hostilities between Serbia and the Austro-Hungarian Empire, pulling in other European powers as a consequence of existing alliances and rivalries. That is to say, the familiar development of historical analysis, offering a more nuanced account of longer-term underlying causes rather than focusing solely on short-term political events and chance occurrences (‘proximate’ causes), is based on a counterfactual
argument: would the First World War still have broken out, albeit a little later, if Ferdinand had not been shot that day? The standard answer, given the familiar structural features and dynamics of early twentieth century European politics, is that it would have done.

It is clear, therefore, that some form of counterfactual thinking is not alien even to conventional history, but is rather fully integrated and taken for granted within it – but it could be explored more explicitly and in more depth in order to improve the quality of the analysis of causation, and hence improve our understanding of the past. Above all, counterfactualism offers a means of preventing or limiting ahistorical error, above all insufficiently interrogated assumptions about inevitability, and the tyranny of hindsight (Kaye 2010). It can certainly be an advantage for the interpretation of the past to know what happened next, simply because it builds up a knowledge base of what possible causes may have what possible consequences, but there is a substantial danger of assuming that a specific type of cause can only have one specified consequence, or that in a given situation only one outcome was likely. That things did happen this way should not expunge uncertainty about whether they had to happen this way; an explicit focus on counterfactuals can restore a sense of the openness of the past, of questions of contingency and necessity, and of the idea that what did in fact happen was not necessarily the most probable outcome, simply because it is the one that occurred. Historical events are always over-determined, with multiple causes, and never sufficiently isolated for deterministic description (Tucker 1997); how else can we determine the significance of different conditions except by reference to events and alternatives?

**Constructing and Evaluating Counterfactuals**

If the necessity as well as the utility of counterfactualism is accepted, the next step is to develop them in the most effective and persuasive manner possible. The first important question is how to identify the counterfactuals that are most worth investigating in a given context. Conventionally, attempts at counterfactual history have concentrated on what appear to be obvious ‘turning points’, moments when things could clearly have gone differently: military engagements, most obviously, but also individual decisions made by those in power. It may be argued that it is sufficient for us to be able to imagine alternative possibilities to what in fact happened, but a stronger argument (e.g. Ferguson 1997) is that we need contemporary evidence that the alternative was considered at the time.

Certainly we can identify this issue in relation to the decisions of individual historical actors; the fact that, in retrospect, the actions of a monarch or general appear critical in determining the subsequent course of events is insufficient to demonstrate that a different outcome was possible, in the absence of evidence that the individual could have chosen differently. This evidence might be provided by documents such as the records of meetings or the private papers of the individual concerned, or (especially in earlier, less well documented periods) it could be a more tentative synthesis of what is known about their general character and/or the climate of thought at the time – the existence of evidence, even if not directly related to the decision-maker, that certain ideas were conceived and capable of being held by such people.

In the outbreak of the Peloponnesian War between Athens and Sparta, there are no conceivable grounds on which to claim that the Spartans might have pursued a policy of pacifism, as that would have been entirely anachronistic; but there is a range of evidence attesting to their reluctance to embark on large-scale hostilities or to see Athens as a real threat to their interests.
Thucydides’ account shows the process whereby they were persuaded by their allies that war was necessary (highlighting, of course, that Thucydides did not regard the war as inevitable in any straightforward manner); on that basis, it is easy to construct and justify a counterfactual in which war was at least postponed for some years, and to consider the consequences of this for the balance of power between the two states and their allies.

In part, the choice of counterfactuals depends on what questions we wish to explore. It may be legitimate and useful to consider how far a decision or a battle was a foregone conclusion, exploring the evidence for the possibility of alternative outcomes (indications that a range of options was considered by those making the decision, evidence of specific circumstances related to a military engagement that might have led to a different outcome result – messages not being transmitted, reinforcements failing to arrive, chance events like a lucky arrow hitting the king in the eye). It can be equally legitimate and useful to consider whether the outcome of such a decision or battle was a foregone conclusion, in which case the counterfactual is taken as a starting-point for discussion rather than as the object of analysis. The first approach seeks to explore the significance of conditions in establishing the range of possible outcomes and their relative likelihood, the second examines the chain of causation arising from that moment.

Counterfactual history is highly responsive to, if not dependent on, mainstream historical narratives and their identification of turning points and inevitable developments; this is an obvious reason why it has tended in practice to focus on battles and political decisions, and why the claim is sometimes made that only documented alternatives should be considered (anything else strays too close to fiction, and to the conventional stricture against historians inventing things unnecessarily). However, it is clearly possible if not probable that certain possible outcomes may be visible only in retrospect and – most controversially – on the basis of modern theoretical understanding. This is especially relevant when we move away from conventional forms of narrative history to the study of longer-term developments in economy, society, and culture, which were largely or entirely invisible to contemporaries.

For example, it is well known that the classical political economists in the late 18th and early 19th centuries failed to recognise the revolutionary impact of a switch to a mineral-based energy economy (coal and other fossil fuels in preference to wood and muscle as sources of power), and so continued to assume, on theoretical grounds, that there were strict physical limits to the possibilities for growth in productivity, even as the English economy was setting about proving them wrong (Wrigley 1988). It is only in retrospect, with the benefit not only of more copious data but also a better understanding of the processes of economic change, that we can see 18th-century developments as a critical turning point in the emergence of the modern economy, and hence explore questions about the relative significance of different conditions that shaped the process and the consequences if it had not happened. (We can also recognise that this account of the Industrial Revolution as a revolution in energy use is already derived from an implicitly counterfactual analysis, moving away from a narrative focused on mechanical inventions (the spinning jenny, the steam engine) to explore the conditions under which such things could be invented and/or have the effects that they did.)

Modern social-scientific theory, and modern scientific knowledge, can identify possibilities that contemporaries were wholly unaware of; part of understanding the impact of the Black Death on late medieval Europe, for example, rests on our having a superior knowledge of *Yersinia pestis* and
the vectors of its transmission, so that we can develop a sense of the range of possibilities – some more apocalyptical, some much less so – according to different models of the spread of the infection and its impact on different populations (cf. Hawthorn 1991). This can offer an important corrective to the traditional historical narratives, emphasising that what appeared significant to contemporaries (and to many of their successors) might be far less important in the longer term. Indeed, one might apply counterfactual arguments to events that are apparently insignificant, precisely in order to open up the question of whether our conventional accounts of the past are as inevitable or significant as they tend to appear to us, and/or to explore our assumptions about importance and priority.

The obvious risk is that these accounts immediately appear more speculative, resting either on confidence in modern theory as a true and universally applicable analysis of the world (the claim that we understand the medical aspects of the Black Death being more widely accepted than the claim that we understand human psychology regardless of cultural difference) or on the historian’s imagination. This leads to the second, and more problematic, question: how are counterfactual narratives to be evaluated, given that we cannot have evidence for what did not happen?

Science professes to draw no conclusions but such as are based on matters of fact, things that have actually happened; but how can any amount of assurance that something actually happened give us the least grain of information as to whether another thing might or might not have happened in its place? Only facts can be proved by other facts. With things that are possibilities and not facts, facts have no concern. If we have no other evidence than the evidence of existing facts, the possibility question must remain a mystery never to be cleared up. (James 1956: 151-2)

The conventional answer is that these strictures apply to all historical interpretation, not only to counterfactuals (Ben-Menahem 2016: 378). The criteria that Tetlock and Belkin (1996: 16-31) put forward for evaluating counterfactuals are, for the most part, those which would be used to evaluate any historical account: well-specified antecedents and consequents; logical consistency; ‘minimal rewrite of history’, which could also be labelled ‘historical consistency’, compatibility with a reasonable interpretation of the existing evidence and limits on how much can reasonably be invented; reliance on ‘strong theory’; and finally ‘projectability’, which is the one criterion specific to counterfactualism, and will be discussed further below.

Logical consistency and clear specification of the argument are obvious, incontrovertible virtues; reliance on ‘strong theory’ is more controversial among historians, and raises the question of which strong theory is adopted – but these are familiar issues in historical interpretation. The idea of ‘historical consistency’ is most interesting and important. It suggests that, if we explore the world that would have resulted if only this one change occurred – the effect of minor changes in starting conditions, as Ben-Menahem (2016) has argued – that we can continue to make use of existing historical data for the period in order to discern the likelihood of different developments and then evaluate them.

It is generally a safe assumption, over time, that initial change would have begun to have wider consequences, so that the counterfactual world will deviate ever further from the world as we know it; hence the further the counterfactual narrative is extended through time, the more speculative it becomes. A classic example is the economic historian Robert Fogel’s analysis of
whether railroads were essential to 19th-century American economic development, which begins with solid data about the relative costs and carrying capacities of canals compared with railroads, and the requirements of the economy, to argue that the historical dominance of railroads was partly contingent and only gradually led to path dependency (Fogel 1964). As the narrative progresses, however, it begins to speculate about the likelihood of the motor car being developed decades earlier than it actually was, as a result of the absence of railroads, an idea that is impossible to test usefully.

Underlying this argument is the philosophical question, raised by David Lewis (1973): how much would the world already have to be different in order for the counterfactual to have been possible? We cannot simply imagine a world in which kangaroos have no tails but everything else is the same; other things must have been different for this one thing also to be different. This resembles the criticism raised against counterfactual history by E.H. Carr: if a particular cause [e.g. a battle] led to the ruin of a state, then there must have been a general cause which meant a state could be ruined by a single battle, rather than everything being reduced to that moment (Carr 1964: 101; cf. Hawthorn 1991: 8-9). This then implies that the counterfactual narrative must begin in the past, rather than at the moment of obvious divergence from the familiar account, for such a divergence to be plausible – and hence the historical data from that moment of divergence will already potentially be different, rather than offering an initially firm basis for the counterfactual. This raises the spectre of an endless regression, in which one has to develop counterfactual narratives of how the data might have been affected by earlier changes, in order to draw on them in evaluating the main counterfactual. But in practice, for shorter-term counterfactuals (or for longer-term ones which focus on slower, longer-term changes in economic, social or demographic structures), the evidence offers a good-enough approximation.

Once again, the seriousness of this objection depends on the aim of the exercise; the evidence problem would be more significant if our aim were to argue strongly that the alternative narrative is how things would have turned out with different starting conditions, than if we simply seek to suggest that they might have turned out this way. The aim of past-focused counterfactuals is not so much to develop alternative histories as an end in themselves – that is something primarily for the writers of speculative fiction – but rather to explore alternative pasts, and above all to evaluate the conditions under which these alternatives could reasonably be considered to have been possible or even likely, as a means of alerting ourselves to contingency and possibility, and to our habit of assuming that the present is the only possible world and hence is the measure of the past. Counterfactual history not only develops out of a reaction against conventional narratives, its primary function is to improve our understanding of the past through a critique of those conventional narratives.

Present-Focused Counterfactuals

A second approach to counterfactualism is the attempt at refining and deepening our knowledge of the world in general, by using counterfactuals as a means of testing normative theories (e.g. Lebow 2010). Rather than a modern social-scientific theory serving as the or a means of legitimising a counterfactual narrative of past events, as discussed above, the database of past events becomes a means of evaluating such theories – or, more concretely, of considering the circumstances in which their normal assumptions may not apply, and the impact on them of small changes in starting conditions.
Theories generate their own counterfactuals; the general principle, for example, that a ‘rising’ and a ‘ruling’ power will always come into conflict raises the question of whether this is invariable, and, if not, what the significant reasons are why it would not apply in a specific case (often leading, as in the case of Allison’s Thucydides’s Trap, to a more qualified statement of the principle). Assumptions about human motivation (e.g. that states are basically rational and driven by concerns with honour, interest and fear, as stated in standard Realist approaches to international relations) can be tested against specific examples and, more importantly, tested against an evaluation of the range of possible outcomes, rather than assuming that the actual historical outcome was the only possible or probable one (whether or not this conformed to the predictions of the theory in question). The main effect of such an approach is not to undermine theories altogether, given that in most cases a reasonable explanation can be put forward to explain why the theory did not hold in a particular situation, but to limit their claims to universality. Many theories of human society express themselves in terms of ‘all other things being equal’; counterfactual approaches seek to explore in explicit and well-articulated terms how often, and in what circumstances, all other things may indeed be equal.

As discussed above in relation to normative theory more generally, the difficulty of such an endeavour is that there is no fixed, objective account of the past against which theoretical predictions can be measured; other descriptions and narratives are always possible, and the data is rarely if ever adequate for falsification (Weber 1996). The more limited the theory, both in terms of its focus (a closed rather than open system) and its chronological extension (covering recent, well-documented events rather than the more distant past), the more plausible such a test against the historical data is likely to be; a counterfactual about the impact of a given economic policy over the short run, for example. With more ambitious and wide-ranging theories, such as a variant of the Thucydides’s Trap model, evaluation both of the model and of possible counterfactuals (could the confrontation of the United States and the Soviet Union during the Cold War in fact have resulted in open conflict? under what circumstances?) becomes rather a matter of competing idiographic accounts, evaluated by their relative logical coherence, plausibility, coherence in relation to the historical record and so forth.

As with past-focused counterfactuals, which are most effective as a form of critique of conventional historical narratives and their common fallacies of assuming inevitability and necessity, counterfactual analysis of contemporary normative theories serves as a vaccine against determinism and excessive simplification. In particular, examining alternative possibilities and evaluating the likelihood of the negative case counters the innate human tendency towards confirmation bias, namely selecting and privileging evidence that supports the hypothesis at the expense of other, less positive, evidence, hence developing over-confidence in one’s knowledge and understanding (see e.g. Baron 2000). For example, various reasons might lead an observer of modern global politics to concentrate on the rivalry between the United States (as the undisputed ‘ruling power’ since the Second World War) and China, not least the fact that the latter’s rise in recent decades has been so spectacular; but we need to consider explicitly how we can determine whether we are now living in a bipolar world, as the Thucydides’s Trap model assumes, or whether multipolarity offers a better basis for understanding (as e.g. Wright 2017 argues).

This inevitably reduces the explanatory and predictive power of theory by highlighting the possibility that other things might have occurred, that conditions could have been different, and
that human, like historical, situations are invariably open and over-determined, with multiple interdependent factors shaping developments, rather than the simplified and linear world of most models. This insight is not necessarily welcomed by many social scientists, and it remains unclear whether a counterfactual approach can lead to improved theories rather than simply to scepticism about theory; clearly, however, it can be taken as a vital hedge against over-confidence in our understanding of the world and ability to make reliable forecasts.

**Future-Focused Counterfactuals**

Thirdly, there are counterfactuals that are focused on the future. The majority of these make use of past- and present-focused counterfactuals and seek to project them forward. As mentioned above, Tetlock and Belkin offer ‘projectability’ as one of their criteria for evaluating attempts at developing counterfactuals: “The same causal principles that allow us to retrodict the past should allow us to predict the future” (30-1). However, it is not clear that this approach differs significantly from the use of normative theory, whether or not refined through counterfactual thinking, discussed in the previous section. It assumes either the continuing validity of the theory, and by implication a continuation of existing conditions into the future – a more persuasive claim in the case of physical laws than social scientific ones – or the objective validity of an analogy between the present situation and a past case or cases which then indicate the range of possibilities for future outcomes. The primary goal of Allison’s Thucydides’s Trap model, for example, is to highlight the possibility, but not inevitability, of escalating conflict between the United States and China, on the grounds that their present relationship strongly resembles historical situations in which conflict has proved the most likely result. However, the fact that, supposedly, conflict resulted in such situations more often than not is not in fact evidence that it was or is an inherently more likely outcome in such situations, even if one accepts that there these are persuasive interpretations of all the different historical case studies and that there are persuasive analogies to be drawn with the present.

Secondly, we find certain attempts at writing ‘counterfactual’ accounts of the future, justified on the grounds that the techniques of constructing counterfactual history are directly relevant to such an exercise – counterfactual accounts are speculative histories of “a ‘future’ that did not really happen”, from the perspective of the ‘turning point’, even if from our perspective they are equally set in the past (Weber 1996: 277). In what sense are such accounts counterfactual, since they relate to something that might still happen, rather than representing an imagined alternative to what did happen? Weber sets the counterfactual approach against what he terms “official” futures, which may be understood as either the unthinking assumption of continuity (or at least the assumption that changes will be only superficial rather than structural) or the narratives currently preferred by those in authority (governments, think tanks, mainstream media), which tend to limit the thinkable possibilities and for the most part emphasise a single line of development. The aim in this exercise is therefore not primarily to offer a reliable prediction of what the future will be like, but to open up questions about the range of possibilities – especially given that humans appear to find it difficult to engage with ideas of non-linear change or conceive of how far the world might alter around them.

It may be that the human mind cannot anticipate in a reliably predictive sense nonlinear change, but it is certainly possible to ask ourselves probing questions about the possibility. Developing studies about how an idea could take off, and what the world would be like if
it did, is one way to think about preparing individuals, states, or any social collectivity for futures that might happen. (Weber 1996: 275)

Such futures are intended to be provocative; to question present assumptions and confidence, to emphasise the extent of uncertainty, to direct attention to the possibility that current developments may have unexpected results and that world-changing developments may emanate from things which are currently beyond out view. We might consider a book like Peter Frase’s *Four Futures: life after capitalism* (2016) as an example of such an approach, since its most striking feature is the delineation of four different lines of possible future development, rather than a single vision of things to come. Frase identifies key challenges and developments in the present (environmental degradation and climate change, automation, inequality), and then imagines the likely consequences of different possible human responses to these problems. The claim is not that any of these scenarios is necessarily realistic, or more probable (though some are acknowledged to be more utopian than others); rather, they are intended (much like Weberian ideal types) to delineate the outer bounds of possibility, with the probable course of future development falling between these extremes but tending to resemble one rather than the others. In other words, Frase’s analysis rests on developing multiple counterfactual narratives, to be explored in comparison with one another, to expand our sense of what might happen.

The objection to such speculation that the class of things which might happen is effectively infinite, and that claims about their possibility are unfalsifiable, is largely beside the point. Firstly, the aim is not to promote a single future narrative but rather to open up the discussion and indicate that there is more uncertainty than may be generally realised. Secondly, it is possible to argue, using the same criteria that are applied to past-focused counterfactuals, that some future narratives are more plausible than others, even if this cannot be proven. Thirdly, we might look to counterfactuals as an exercise in persuasion and discourse, in telling different stories about the future as we do about the past – and it is the comparison of these different stories, offering different accounts of possible futures, that allows us to engage seriously with the present as a preparation for the ever-emerging future (cf. Weber 1996).

**Counterfactuals and Anticipation**

Counterfactual thinking about the future is not scientific, or objective, or empirically robust. It often depends heavily on the assumption that past and present offer at least some guide to the future; that we can identify immanent tendencies in the present and explore their possible consequences, and that our existing data sets are our most important resource for such an exercise. It can indicate possibilities, but its main utility lies in questioning the more confident and dogmatic assertions of other forecasters; in highlighting the fact that past developments were never necessary or preordained, and in emphasising uncertainty and indeterminacy.

Anticipation is not about forecasting; indeed, it is less concerned with the actual future than with how we think about the future and engage with the present as a result. The inability of counterfactualism to aid in the identification of a single or most probable future is therefore not a problem from this perspective; rather, it directs our attention towards how humans think about the relationship between past, present and future – and how this may lead them into trouble. It offers a means of identifying and correcting common cognitive biases: confirmation bias, framing, reliance on unexamined heuristics such as analogies, and, above all, hindsight bias (Kahnemann, Slovic & Tversky 1982). If ‘shoulda woulda coulda’ are indeed the last words of a
foil, that doesn’t refute the idea that many people nevertheless evaluate present situations and future possibilities by considering what might have been in the past.

This is precisely the aspect of Thucydides’ account of the war between Athens and Sparta that Allison’s model ignores; far from offering a single explanation of a deterministic cause, the narrative highlights both the complex interaction of multiple causes (individual actions and decisions, the characters and cultural assumptions of different states, and the continuing consequences of past events, as well as structural factors) and the multiple ways in which things might have turned out differently (Hawthorn 2014; Tordoff 2014). The same is true of Thucydides’ account of the subsequent course of the war, which is presented as neither predetermined nor purely contingent, but always open, even if there was a recognisable tendency towards path dependency. Above all, the ‘lesson’ which Thucydides’ readers can draw from a better understanding of these past events is that human beings suffer from multiple failings in evaluating the present and anticipating the future: oscillation between over-optimism and excessive pessimism, assumption of greater knowledge and confidence than is rationally justified, susceptibility to false analogies and confirmation bias, and a habit of reinterpreting events afterwards in a self-justifying manner rather than learning properly about their own errors.

The primary contribution of counterfactual thinking to anticipation is therefore psychological, cultural and ethical, as Audra Mitchell has argued: “It cannot give us predictions or certainty, and it can’t prove that everything will be ok, or tell us how to ensure this. But it can help us to see possibilities, to scope the boundaries of our knowledge, to appreciate the limits of our agency and to expand our ethical sensibilities” (2013). Mitchell’s argument focuses on the specific issue of the storage of nuclear waste, where it is essential for us to imagine futures that are unknowable or unthinkable – to break through what she calls ‘futural amnesty’ or the forgetting of the future – because what we do or do not do in the present will inevitably have far-reaching consequences. Relatively few issues – climate change is the other obvious one – combine such long time-scales with the knowledge that our actions or inaction will have consequences for the future. But the wider points she makes about counterfactual thinking and anticipation are valid also for shorter-term concerns.

By combatting assumptions of determinism in human history, and emphasising that there are multiple possible futures, counterfactual narratives work against a crippling nihilism and sense of a lack of agency. They can offer a qualified sense of hope, that the future might not be entirely bleak and certainly that its bleakness is not pre-determined – and hence, an ethical obligation towards the people of the future, to seek to build a better rather than a worse future. Conversely, imagining horrific and nightmare possibilities, including human extinction or degradation, provides the ethical imperative for trying to avert such futures. Futural amnesty is a protective strategy, allowing people to get on with their lives without constant awareness of horror (this resembles Friedrich Nietzsche’s arguments about the inability of humans, with rare exceptions like Thucydides, to face too much reality), but it is also an abdication of responsibility for the harm that we might thereby do to unknown future others. Counterfactualism tells us that the future is still open, that we can usefully seek to anticipate it – even as it also highlights the ways in which humans constantly fail in this effort – and that we have an ethical obligation to try.

**Summary**
The qualities that make counterfactualism marginal within historiography are precisely those that put it at the heart of anticipation: its emphasis on the openness of historical processes and the limits of our understanding of them, its critique of determinism and over-confident forecasting, the intellectual and rhetorical techniques it offers for developing and evaluating narratives about things that have not happened, and the ethical obligations it highlights to engage with future possibilities.

References


Ben-Menahem, Y. (2016) If counterfactuals were excluded from historical reasoning… Journal of the Philosophy of History 10, 370-81.


**Keywords**

Counterfactuals, counterfactualism, history, historiography, forecasting, theory, cognitive bias, Thucydides, narrative, normative theory, historicism, future, ethics