

Counterfactual thinking as a scientific method

Wenzlhuemer, Roland

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Wenzlhuemer, R. (2009). Counterfactual thinking as a scientific method. *Historical Social Research*, 34(2), 27-56.
<https://doi.org/10.12759/hsr.34.2009.2.27-56>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:

<https://creativecommons.org/licenses/by/4.0>

Counterfactual Thinking as a Scientific Method

Roland Wenzlhuemer *

Abstract: »Kontrafaktisches Denken als wissenschaftliche Methode«. Despite their alleged uselessness, we frequently encounter instances of counterfactual thinking in everyday-life situations. During the last two decades, psychologists have examined this phenomenon and have been able to show that counterfactual thoughts about past (and, therefore, unchangeable) events can be useful for the thinker. This article retraces the effects of counterfactual thinking in everyday life and seeks to translate them into an academic context. Can counterfactuals produce similar analytical benefits in the sciences? And if so, what exactly are those benefits? After briefly examining the psychological aspects of counterfactual thinking, the article discusses its potential uses from a general scientific perspective before using counterfactual history as a concrete example. Eventually and for good measure, attention is drawn to the clear and present dangers associated with counterfactual thinking in an academic context.

Keywords: counterfactual thinking, counterfactual history, contrast effect, hindsight bias, self-protective bias.

Introduction – The Psychology of Counterfactual Thinking

If ‘ifs’ and ‘ands’ were pots and pans, there’d be no work for tinkers.” – “Wenn das Wörtchen ‘wenn’ nicht wär’...” – “Avec des ‘si’ et des ‘mais’, on mettrait Paris dans une bouteille.” – “Če čebula ne bi imela če, bi bila bula. (Popular Wisdoms in English, German, French and Czech).

On 20 August 2008 Spanair Flight JK 5022 crashed at Madrid Airport immediately after take-off. More than 150 people died in the crash and the subsequent fire. In the following days, media coverage was extensive. Experts seemed puzzled by the accident and offered a host of different explanations. Soon there were allegations that the crew had kept anxious passengers from getting off the plane after an aborted departure. Relatives of the casualties raised all sorts of accusations and demanded justice. And yet within all this tragedy and chaos, among the most impressive and unforgettable pictures were those of the couple who had arrived three minutes too late at check-in and had not been allowed on the plane anymore. Interviewed by Spanish television, they were both in a state

* Address all communications to: Roland Wenzlhuemer, Cluster of Excellence ‘Asia and Europe in a Global Context’, University of Heidelberg, Karl Jaspers Centre; Voßstraße 2, Gebäude 4400, 69115 Heidelberg, Germany; e-mail: wenzlhuemer@asia-europe.uni-heidelberg.de.

of severe shock. She has not been able to say much at all, he briefly stated that they were grateful and would celebrate a second birthday.

This tragic and unfortunately non-fictional event provides one of the classic “what if” scenarios. The couple had been determined to take the flight and only some unforeseen coincidence, mere chance, saved them from a potentially fatal disaster by minutes. It is absolutely reasonable to assume that the lucky couple will for a long time be haunted by questions starting with “what if” or “what might have been”. Two factors are likely to trigger such thoughts in this particular case: the magnitude of what was at stake and the closeness of the alternative. While even in this extreme example there seems to be no practical use in thinking about alternative outcomes and potential fatalities, the closeness of the event – i.e. the extraordinarily narrow escape by sheer chance – practically forces such thoughts and questions on the subjects’ minds. Such an “imagination of alternatives to reality is called *counterfactual thinking*” (Markman et al. 1993, 88). A different future was so close at hand that there is no way but to contemplate this alternative. It seems that this closeness somehow highlighted the existence of an open future, of other possibilities, of contingency, when usually we tend to assess the *status quo* as given, as the necessary and only possible outcome of a situation.

The psychologist Neal Roese emphasizes the role of closeness as an activator of counterfactual thinking. He says that in particular *downward counterfactuals* (i.e. when the imagined alternative is worse than the actual situation) are rarely activated spontaneously and usually require a trigger. Outcome closeness can be such a trigger (Roese 1997, 136-7). Roese also refers to the classic (and in this case fictional) example of a missed and later crashed plane. In addition, however, he cites a study that tested the role of closeness in counterfactual thought activation. In the fictional scenario, a man had forgotten to submit his insurance policy either three days or six months before a fire broke out. In a subsequent thought listing¹, counterfactual thoughts about (not) submitting the policy appeared more frequently in the closer scenario in which three days had made the crucial difference (Roese 1997, 137).

In both examples closeness (together with the seriousness of the situation) clearly is the principal activator of counterfactual thinking. Its purpose, however, differs widely. One could assume that there is no use crying over spilt milk in the fictional example. The man was not insured and the insurance company, therefore, will not carry the costs caused by the fire. However, as a direct action (or inaction) of the thinker has actually brought about the unfavourable situation (not the fire but the financial loss), thinking about counterfactual scenarios can highlight causal relations in this case. If the protagonist had been more diligent and submitted the insurance policy earlier, the insurance com-

¹ The thought-listing technique is a method in psychological research in which a subject is asked to list everything that he or she is or was thinking.

pany would now cover the damage. To the counterfactual thinker a clear causal relation between his action (or inaction) and the present situation becomes visible. For the future this opens the prospect of acting differently next time, of learning from past blunders. Neal Roese has called this the *causal inference effect*. With the help of counterfactual thought experiments, alternative actions and outcomes are simulated, causal relations can be established and, eventually, behavioural patterns can be adjusted (Roese 1997, 141).

In the non-fictional example, no conscious action of the protagonists was involved in avoiding the situation and, therefore, counterfactual thoughts are seemingly pointless. By a sheer stroke of luck, the couple has been saved from almost certain death. The plane had not been missed on purpose. There are no causal chains involved in the example that could have been altered by conscious actions of the protagonists. The sole purpose of the recurring counterfactual thoughts is to contemplate alternatives, other possible futures and, therefore, big questions such as the role of fate, predestination, contingency, God or such like. And yet, counterfactual thinking can have a very important (if not always welcome) effect in such a situation as well. When the thinkers compare the actual result or situation with the alternative “what might have been”-result, this also leads to a reassessment of the actual situation in the light of the alternative outcome. Roese calls this the *contrast effect*. The realized situation can appear worse to the thinker than it actually is, if the counterfactual alternative would have been much better. On the other hand, an outcome can seem better than it is, if the alternative would have been worse. According to Roese, comparisons with *upward counterfactuals* are much more common than ones with *downward counterfactuals* and, therefore, the *contrast effect* is rarely beneficial for the thinker (Roese 1997, 140-1). The non-fictional example about the missed plane might, however, constitute a rare exception in this respect.

Keeping the *contrast effect* in mind, we can safely say that a situation derives at least a part of its actual meaning from the alternatives that would have been possible at any given time. Therefore, counterfactual thinking is a fundamental component of assessing a situation and one’s satisfaction with a particular outcome. Interestingly, Roese has also demonstrated that there is a strong link between the *contrast effect* and the *causal inference effect* in counterfactual scenarios with a potential learning curve. Protagonists will usually not feel as bad about a situation with an *upward counterfactual* (e.g. a failed exam), when a causal relation (e.g. heed the supervisor’s suggestions) has been discovered that promises better results for the future (Roese 1997, 141, Markman et al. 1993). The prospect of learning from one’s mistakes, therefore, seems to make a past blunder more acceptable.

Counterfactual Thinking as a Scientific Method

Process, n. [...] 8. A continuous and regular action or succession of actions occurring or performed in a definite manner, and having a particular result or outcome; a sustained operation or series of operations. (Now the most common use.). (Oxford English Dictionary).

What we can learn from psychological studies on counterfactual thinking is that contrary to popular opinion there seems to be some use in crying over spilt milk. Even if this might not be the thinker's intention, counterfactual thinking has clearly measurable effects in everyday life – some of them negative, some of them positive. It is subconsciously deployed as an analytical tool to unearth causal relations or compare a situation against potential alternatives, when a re-run of the process is not possible. Researchers in the social sciences and humanities intimately know the problem that their objects of study only rarely lend themselves to re-runs and experiments. Therefore, the question arises whether there is any chance of intentionally applying similar counterfactual methods and strategies in order to examine research questions in a scientific context. After all the general purposes of analytical thinking in everyday life and in the social sciences and humanities are closely related: to *understand processes* and to *create meaning*. As a matter of course, complex processes can only be understood with a secure grasp of the causal relations within. Similarly a situation or action can only have a clear meaning in contrast to other potential courses of action. Therefore, these are two fields in which counterfactual methods can be very valuable from an analytical perspective.

There are, however, at least two more potentially beneficial effects that counterfactual thinking can have in an analytical context. While I acknowledge that both might seem somewhat trivial at first glance, I am convinced that their potential – and, indeed, their necessity – can hardly be overestimated. First, counterfactual thoughts can help researchers to overcome traditional, established boundaries and pathways of enquiry. They can allow us to zoom out, re-focus and contemplate a fresh perspective, when in danger of becoming professionally blinkered. And second, counterfactual thinking can be very informative regarding the thinker's very own positions and viewpoints. It can be a means to create self-awareness in researchers regarding their own perspectives and potential biases. Taken together I see four main areas – causation, contrast, *Spielraum* and self-awareness – to which counterfactual thinking can make worthwhile contributions in an academic setting. In the following, I will briefly elaborate on these four areas from a more general viewpoint before then using counterfactual history as a concrete example from the disciplines.

Causation

The search for causal relations lies at the very heart of modern scientific enquiry and unites the most diverse disciplines – from the natural sciences via

medicine to the social sciences and the humanities. Most research in all these fields is concerned with the analysis and eventually with the understanding of processes of some sort. While, of course, the object of study itself will be static or unchanging in many cases, it will usually be part of a larger process. Understanding the status of the object requires understanding its position in the process in question – its preconditions and its consequences, its predecessors and its successors. Identifying and eventually understanding causal relations is absolutely indispensable in this regard. While united by the search for causation, the different disciplines have devised and perfected very different methods to conduct their studies. In the classic natural sciences – such as physics or chemistry – the controlled experiment provides the principal tool to test hypotheses and identify causal relations. In the process or system under scrutiny individual factors will be changed by the researcher. By comparing the subsequent results in the original system and in the manipulated system, causal factors and chains can be isolated. Depending on the researcher's prior knowledge about the object of study, controlled experiments can range from very informed and precise enquiries to painstaking trial-and-error endeavours. Eventually, when a series of experiments has generated a large enough data basis, extrapolations from these data can be made. Scientific theories and laws can be devised from the data and will later then be tested further, refined, adjusted and someday maybe invalidated.

In the social sciences and especially in the humanities many objects of study do not lend themselves to being examined in controlled experiments. While there are certainly research questions – for instance in sociology or communication studies – that can be answered with the help of controlled experiments, this path is barred for other disciplines such as history, literature, philosophy, law, cultural studies and even economics to name but a few. Here, the objects of study cannot (or should not) be willingly manipulated. Experimental re-runs are either impossible or would not produce usefully results – or both. For instance, a historical event cannot be re-run under controlled conditions, an widely circulated piece of literature cannot be unwritten or taken out of circulation, the grammar of a widely-spoken language cannot be changed on purpose and existing (and functioning) law should not be altered without exactly knowing one's trade. Therefore, these disciplines have developed a different set of methods that usually rests on the careful observation of the object of study in as many different settings as possible – without, however, intentionally manipulating individual factors.

Causal relations are established by diligently observing the course and rationale of particular processes from different perspectives and with different tools and methods. While there usually is no explicit control group (like it would be the case in controlled experiments), the research process implicitly contains a control element – counterfactual assumptions about the object of study. This is to say that every interpretation of the observed data automatically

implies a counterfactual element. In most cases, investigators in the social sciences and humanities lack the particular form of “hard” evidence that much research in the natural sciences rests on. Their evidence comes from diligent observation and interpretation – an attempt to make sense of the different factors and relations in a particular process without having controlled access to it. In the investigator’s mind, such interpretation always automatically includes a hypothetical scenario in which a factor, indeed, is manipulated. Even if the contemplation of such hypothetical counterfactual scenarios is rarely explicitly acknowledged, it is part and parcel of the research process.

To provide a very simple example, let us assume that we want to understand why student X failed the history exam. In order to do so we have to place the event of failing the exam in a wider process that led to the event (likewise we should also look at the consequences which in this particular example, however, are not of interest to us). Why did student X fail? One possible explanation could be that he/she did not learn enough. But there could be other causes as well: maybe the exam was too hard or even unfair; maybe student X had a headache and could not concentrate; or any other reason one can think of. Obviously we cannot re-run the process under controlled conditions. We do not have the option of conducting a controlled experiment in order to establish clear causality in this case. We will, therefore, carefully observe and analyse the process of failing the exam, we will compare it with other failed exams and then also compare it to passed exams until some causal pattern becomes visible. Let us assume that the pattern, indeed, directs us to the hypothesis that students who learn diligently pass exams and others usually do not. We will then – often unconsciously test this hypothesis by asking ourselves what would have happened if student X had actually learned diligently. We will re-run the process under scrutiny in our minds in order to create a control group. “Thus, running a counterfactual simulation in one’s head constitutes a proxy experiment” (Roese 1997, 141).

To my understanding of analytical processes such counterfactual re-runs are an omnipresent and indispensable part of our work whenever we try to understand causal relations. The only way to test whether an assumed cause c does actually have the effect e is to manipulate c and drop it from the equation – even if this can only be done in our minds. I am convinced that such counterfactual thinking – however implicit and unconscious it may be – forms a crucial part of every enquiry into causality and replaces the controlled experiment. Assuming that this is, indeed, correct one critical question immediately arises: if we implicitly apply counterfactual thoughts in every analysis, why should we need a more explicit acknowledgement thereof? If we are doing it anyway and always, what can be gained by becoming more conscious about the counterfactuality of much of our thinking? Or in brief, if we have been doing this successfully for ages, why should we now start to think about it explicitly? A first answer can already be found in the philosopher David Lewis’ seminal treatise

on causation from the year 1973. Lewis names a number of problems that can arise when trying to isolate causal relations in a system: “If c belongs to a minimal set of conditions jointly sufficient for e , given the laws, then c may well be a genuine cause of e . But c might rather be an effect of e : one which could not, given the laws and some of the actual circumstances, have occurred otherwise than by being caused by e . Or c might be an epiphenomenon of the causal history of e : a more or less inefficacious effect of some genuine cause of e . Or c might be a preempted potential cause of e : something that did not cause e , but that would have done so in the absence of whatever really did cause e ” (Lewis 1973, 556-7).

In statistical terms, we could also say that correlation does not imply causation. In uncontrolled research scenarios, counterfactual thinking can help us to avoid confusing correlations with causal relations by hypothetically manipulating the assumed cause c . While we are, as put forward above, automatically applying counterfactual thoughts in such situations, only a deliberate and explicit adoption of this mode of thinking will allow us to steer clear of the pitfalls that Lewis identified. Intentional counterfactual thinking can lead us deeper into the complexity of a studied system of relations than implicit thoughts will be able to. A simple example can be found in an article by Joseph Halpern and Judea Pearl who in turn have adapted it from other writers in the field. Two fires (A and B) are advancing towards a house. Fire A is a bit faster and burns down the house before fire B can do so. Typically, fire A will be identified as the “actual cause” for the devastation. This example challenges classic counterfactual definitions of causation – such as Lewis’ – as the house would have burnt down (due to fire B) even if fire A had been extinguished (Halpern and Pearl 2005, 843-4). In research scenarios that are not as deliberately obvious as the example of the two fires, the interpretation of such more complicated interrelations of causes and effects are beyond the capacity of implicit counterfactual thinking. In the social sciences and humanities, most objects of study encompass an often confounding multitude of interrelated factors that, indeed, invites misinterpretation along the lines sketched out by Lewis in the quotation above. An explicit and interest-guided application of counterfactuals in such research scenarios can help investigators to steer clear of Lewis’ pitfalls. Or as the political scientist Richard Lebow puts it: “I use counterfactuals to probe the limits of theory and to develop better means of understanding causation in an open-ended, non-linear highly contingent world” (Lebow 2009, 66).

Contrast

Both the social sciences and the humanities are not only concerned with understanding causal relations and the inner workings of a process, they are equally (sometimes even more) interested in understanding the meaning of events,

social practices or artefacts – and especially in the creation of such meaning. While this focus has been made explicit in the so-called cultural turns of the last forty years, it has implicitly informed research interests before as well. The social sciences and humanities are primarily concerned with human beings and their actions, interactions and products. And these in turn are massively shaped by the perceptions, the views, the knowledge of the people (or groups of people) under scrutiny. A historical event or a social practice, for instance, might well be objectively described by the historian or the ethnologist, but a good part of its power, of its meaning depends not on how it actually took place (in a Rankean sense) but rather on how it is perceived, interpreted and remembered by the people.

As mentioned above, the closeness of an avoided (or missed) alternative is an essential activator of counterfactual thoughts in everyday life situations (Roese 1997, 140-1). Accordingly, it is reasonable to assume that similar mechanisms have to be expected whenever alternative outcomes in a scenario under scientific scrutiny have – at one point – been very close at hand. The researcher will need to take into account that in his/her process of study the protagonists' opinions and actions might be influenced by their knowledge and assessment of potential alternatives. This is especially important in all those disciplines that are concerned with the actual perceptions and actions of people – for instance in history, ethnology, law, sociology and many more. Let me refer to a classic historical example in order to illustrate what I mean. Until its final stages, the outcome of the Second World War was largely unpredictable. To many contemporaries it even seemed very reasonable to expect an Axis victory for some time as things were not altogether going well for the Allies. No matter how specialist historians might in retrospect assess German chances to emerge victorious from the war, it can safely be said that for most contemporaries the Allied victory was – all in all – a close one that could not have been surely expected. Beyond any doubt the closeness of an alternative scenario has triggered a lot of counterfactual thoughts revolving around a different outcome of the Second World War. The abundance of alternate history novels playing with the very possibility stands testimony to this.

Together with many other factors – for instance the duration, intensity and particular fierceness of the war – this perceived closeness of an alternative (which, in addition, would have looked very different than the actual outcome) impacts on the symbolic character and the historical and cultural meaning of the war. The perception of the Allied victory as a historical watershed event that shaped the entire course of the remaining century is further increased by imagining the alternative. New meanings are ascribed, old meanings are adjusted. For the researcher this means that the Second World War and its symbolic meaning across the globe can only be understood, if he/she has a certain idea of the alternatives that seemed open (and, indeed, probably) to the protagonists of the study (be it, for instance, Second World War survivors in a

historical study or be it present-day youths in, for instance, a sociological study about far-right extremism today). Here, counterfactual thinking and the explicit invocation of counterfactual scenarios can help the investigator in identifying realistic alternatives and resulting contrasts. Both in the historical and the sociological context – to stick with the examples – the counterfactuals have to be such ones that the protagonists of the study held or hold to be reasonably realistic and influential on the further course of history or on the outcome of some process. In his opening chapter to *Virtual History*, Niall Ferguson suggested that all historical counterfactuals applied in a scientific context must be ones that contemporaries actually contemplated (for which some historical proof should exist) (Ferguson 2003, 86). While I do not support this claim in principle, Ferguson certainly is right as long as counterfactuals are applied to reveal contrasts and the creation of meaning. Both the perceptions of, say, Second World War survivors and present-day neo-fascists can only be influenced by counterfactuals that they themselves view as plausible, even probable, alternatives. Therefore, it must be the first task of the investigator to identify such alternatives correctly. Diligently applied, we can then learn more about ascriptions of meanings and symbolic value to certain events and processes.

‘Spielraum’

Both in order to unearth causal relations in a process as well as in order to invoke counterfactual contrast to learn more about meanings, the investigator will have to follow certain rules in his/her counterfactual analysis. Alternative scenarios should not be deployed randomly. Just like with any other scientific method, there are important guidelines that need to be adhered. For instance, the principles of plausibility and probability of a counterfactual are among the most basic ones (Ferguson 2003, 84-6). To achieve one particular effect of counterfactual thinking, however, the researcher will have to relax or partially suspend some of the rules. I use the German term *Spielraum* here for lack of an adequate word in English. It roughly translates into leeway or elbowroom and denotes the existence of room for manoeuvre – in our case for the investigator. The German term, however, consists of the verb *spielen* (to play) and the noun *Raum* (room or space). Literally it means room to play around and try out different things and ideas. This allusion to playing and experimenting makes the German term a better choice to name the effect. Developing counterfactual scenarios in research can constitute an opportunity for the researcher to think about a problem in unorthodox ways. Here it is not the immediate goal to generate new insights and a better understanding of the object of study. If at all, this would only come in a second, disconnected step. Rather, thinking about counterfactuals allows a researcher to leave the trodden paths of investigation and the established ways of enquiry for some time. It can be used to intellectu-

ally zoom out, playfully interact with the object under scrutiny and create a space in which to let one's mind wander.

To this end, the rules that guide counterfactual analysis in order to infer causation and create contrast mostly do not apply. Here, the counterfactual scenarios invoked by the thinker can range from highly plausible to downright absurd, from very probable to completely impossible. The potential of the counterfactual does not rest in its closeness to the "real world" but rather in the imaginative power it unleashes. Using one's counterfactual *Spielraum* can have at least two beneficial effects. First, it can simply be entertainment and distraction in the – not altogether unheard of – case that a study has reached a temporary dead end. Mostly no usable insights will be generated by such counterfactual thoughts, but their virtue lies in offering fresh (and not always entirely serious) perspectives to the researcher and in thereby relaxing the tenseness of an investigation that does not seem to make progress. This can be a very welcome and necessary contribution. Second, however, letting one's mind wander in such ways might actually produce completely unexpected insights. The history of technology, for instance, knows a fair share of important inventions that were the unintended results of unorthodox, allegedly erroneous investigation. Such serendipitous discoveries are also possible in the disciplines that concern us here – but due to their very nature they can neither be predicted nor expected. More importantly, therefore, new insights will mainly stem from the explanatory power that many *Spielraum* counterfactuals have as well as from the philosophical challenge that they can pose. Many classic thought experiments from the realms of physics or philosophy stand testimony to this. For instance, think of Erwin Schrödinger's famous thought experiment involving decaying (or not decaying) atoms, a flask of poison and a cat that is not to be envied. This artificial and highly counterfactual scenario offered a powerful illustration of a key paradox in quantum physics and made some of the problems of the field better understandable for those not so intimately acquainted with quantum theory. And it opened this new field in physics and its wider implications to discussion in other disciplines such as philosophy. The so-called *Trolley Problem* in ethics can serve as a further example for a counterfactual thought experiment that opens up a formidable intellectual and ethical problem.² Here the artificial character of the scenario creates a situation that is relatively undiluted by external factors and confronts the thinker with a clear-cut problem and no easy ways out. *Schrödinger's Cat* and the *Trolley Problem* both illustrate the explanatory as well as the disturbing potential that counterfactual *Spielraum* in the analysis of a scientific problem can generate.

² The term Trolley Problem refers to a thought experiment in ethics. The setting: a trolley runs along a track to which five people are tied and will soon overrun them. A switch can lead the trolley onto a different track to which only one person is tied. Should the switch be flipped?

Self-awareness

As fourth and final benefit of counterfactual thinking in scientific analysis I would like to refer to the possibility to learn more about one's own views and potential biases as a researcher. If an investigator re-runs a process in his/her mind (and eventually in his/her writing) with a counterfactual point of departure, the counterfactual course of events in the process will also reveal much of his/her views on the object of study. This is especially relevant for the discipline of history and will, therefore, be discussed in some more detail in the following section.

Counterfactual Thinking as a Historical Method

Every action of theirs that seems to them an act of their own free will, is in an historical sense not free at all, but in bondage to the whole cause of previous history. (Tolstoy: War and Peace (BK IX, ch 1)).

As briefly mentioned in the introductory chapter our discussion of counterfactual thinking is based on its broad definition as the “imagination of alternatives to reality” by Markman et al. (1993, 88). The beautiful thing about this short definition is its treatment of counterfactuals as negative functions of reality. On the one hand this allows us to precisely identify and describe instances of counterfactual thinking, while on the other hand a postmodern understanding of the term “reality” provides for a more flexible and inclusive handling of the concept. There are realities for the past, the present and the future. For the past, at least, there might even be a kernel of a common, an objective reality in the Rankean sense on which grew a potentially infinite number of perceived and constructed realities in the postmodern sense of the word. While past realities exist due to the singularity of history, present and future realities emerge from a widely shared belief that there is simply no alternative to certain assumed necessities.³ Even though such present and future realities – that have not yet been fixated by becoming history – can only be established by extrapolation, the widely shared belief in them grants them quasi-definite status. In other words: by looking at past experiences (some of which have been brought about on purpose in the form of experimental studies) we try to identify general laws – be it in physics, sociology or any other field – that will govern present and future processes as well. Therefore, we have always been creating present and future realities by looking into past ones.

Defining counterfactuals as imagined alternatives to reality allows us to include investigative practices that do not focus on explicitly past events in our

³ Such perceptions of unavoidable courses in the present or future are often encapsulated in the catchy abbreviation T. I. N. A. – there is no alternative.

deliberations. Philosophical thought experiments often do not have a clear chronology. In most cases, however, they violate commonly accepted laws or run counter any reasonable estimates of probability. Similarly, counterfactual reasoning in economics mostly does not aim at the past but rather seeks to test alternatives to established economic patterns. To provide one further example, conspiracy theories can also be treated as counterfactual explanatory models of the world. They run counter a widely accepted order of the world, a perceived reality, and offer alternative explanations. All these practices in many different disciplines qualify as methodical counterfactual thinking and can harvest one or more of the potential benefits outlined above by critically exploring their realities of reference.

And yet in many practical research situations, counterfactual thinking will nevertheless revolve around events in the past, about historical processes and situations. This is probably due to the fact that so-called historical realities seem easiest to handle. In most of them, at least a skeleton of generally verifiable events remains. History fixates a set of facts on which counterfactuals can be based. Time cannot be turned back. Not many historians, for instance, will deny that, say, the French Revolution really took place. However, what these revolutionary events mean, how they tie in with a greater historical context, how they were caused and how in turn they impacted on the subsequent course of history, is still – and will probably forever be – the matter of discussions and differing interpretations. It is, I believe, this particular dichotomy between established facts and diverging realities based on the interpretation of these facts that renders historical research a primary field for the application of counterfactual thinking. The facts provide a suitable and easily identifiable point of application, while the existence of differing and often competing interpretations makes the systematic search for causation and meaning in historical processes all the more necessary. This tension inherent to all historical enquiry saw to it that the discipline became one of the first fields of systematically applied counterfactual thinking. Writing as early as in the first century BCE Livy suggests in a famous passage of *Ab Urbe Condita* that Alexander the Great, had he lived longer, would have been defeated by the Romans. This so-called “Alexander digression” has often been referred to as the first account of counterfactual thinking in historiography (Morello 2002). And it already illustrates one of the potential dangers associated with counterfactuals in history – namely that they are conceived in a way that supports the ideological position of the author (in this case Roman superiority). But before we turn to the dangers of counterfactual history, let us for a while contemplate how the potential benefits of counterfactual thinking outlined in the previous section can be applied in historical research.

Causation

Identifying causation in historical processes is essential if we want to understand our past at least to a certain extent. It brings a degree of structural and chronological order into a field that proves to be highly elusive. This is a necessary step if we want to understand, interpret and discern meaning from the past. There would not be much need to intensively study a past in which everything happened without any relation to other events or processes. Nothing could be learned. Nothing would have any meaning. Thus, the search for causation is indispensable – and often very arbitrary at the same time. As pointed out, causation establishes a degree of order in the past that allows us to work with our subject, but it also inflicts an order that often reflects more the researchers' necessities and beliefs than the actual historical pattern of causal relations. Causation can easily be ascribed to or read into something. Identifying causal links is, therefore, both a key trade of the historian as well as a delicate endeavour in which the researcher's own take on history can find a backdoor into the past.

Counterfactual thinking can be one means to identify and test causal relations in a process under scrutiny – in history as well as in most other disciplines. The approach is well-based on the work of philosophers on causation and counterfactual conditionals (for instance, Lewis 1973; Austin 1970; Goodman 1947) as well as on psychological studies (already cited above) that illustrate how counterfactuals can highlight causal relations. Initially, it works along the same principles. In the process under scrutiny, the historian first identifies the potential cause – the event or condition whose role in a causal chain shall be established. This cause is then counterfactually manipulated – just as it is in the philosophical or psychological examples that we have heard of. Then, however, counterfactual history begins to differ essentially from those examples. While the latter usually aim at establishing a binary causal relation between two events or conditions, to the historian cause and consequence cannot be interpreted in isolation and need to be contextualized. In practice this means that in counterfactual history we cannot stop with saying that without a cause c there would not have been an effect e . In itself, this might be a valid and interesting insight. However, unlike many psychological test setups or philosophical thought experiments, counterfactual scenarios in history are not closed systems. History is an incredibly complex system and needs to be treated as such. By manipulating a cause c , not only its relation to the effect e is affected – myriad connections to other events or conditions are altered as well. These changing relations need to be traced as accurately as possible until an alternative timeline unfolds. Only in constructing and arguing for such an alternative course of history, the role of cause c can be properly assessed. This act of construction compels the historian to think much more consciously about direct causal relations between events or conditions than this

is necessary in the de-construction of a simple binary causal relation between cause *c* and effect *e*.

The complexity of the system and the practically incalculable number of possible counterfactual manipulations and subsequent courses of history make certain guidelines for best-practice necessary, if the search for causal relations through counterfactual thinking should be successful. Niall Ferguson claims that counterfactual manipulations should be both plausible and probable. “For example, no sensible person wishes to know what would have happened in 1848 if the entire population of Paris had suddenly sprouted wings, as this is not a *plausible* scenario” (Ferguson 2003, 83). Ferguson then goes even further and says that as a general rule “[w]e should consider as plausible or probable *only those alternatives which we can show on the basis of contemporary evidence that contemporaries actually considered*” (Ferguson 2003, 86). This would mean that only such counterfactual manipulations should actually be made that correspond with scenarios which seemed possible or even likely to contemporaries – for which some form of proof must exist.

Ferguson’s suggestion has many virtues. It considerably narrows down the number of alternative paths and it offers an instantly understandable justification for counterfactual thinking in history as it starts with an assumption that has been contemplated by contemporaries and is, therefore, in a way more historical fact than fiction. The plausibility of an alternative scenario is certainly of the essence, when looking for and testing causal relations. In an implausible alternative (populated with winged Parisians, for instance) the known laws, mechanisms and coherences – with the help of which causal relations will be tested – might be fundamentally different and simply not conceivable. The search for causation needs a plausible alternative scenario governed by the same general laws as the real world. That the alternative has seemed probable (or even possible) to contemporaries, however, has no impact on causation whatsoever and should, therefore, not be considered a general precondition in the selection of counterfactuals – even if we will come back to it in a slightly altered form regarding the contrast inference potential of historical counterfactuals.

Focusing on plausible manipulations only narrows down the number of possible alternatives considerably, but still leaves the investigator with a confounding abundance of possible starting points. The choice between all these possibilities now depends solely on the research interest of the investigator. When a starting point – a so-called point of bifurcation – has been chosen and an alternative course of history is examined, new points of bifurcation will spring up soon. The alternative timeline becomes a “garden of forking paths” (Ferguson 2003, 68-71) through which the researcher has to negotiate by establishing and testing causal relations. As the possible forks and bifurcations increase exponentially with the course of time, in-depth studies of causation are only manageable for short observational periods. Counterfactual history over longer time

spans – such as centuries, for instance – can produce feasible and interesting results, but will be possible on a somewhat cruder and more superficial level only.

Contrast

In his groundbreaking book *Ungeschehene Geschichte*, the German historian Alexander Demandt lists three principal objections that have often been raised against the validity of counterfactual thought in the historical sciences: the meaninglessness of that which has not happened (*das Ungeschehene*), the lack of well-founded research methods and the lack of respect for that which has happened (Demandt 1986, 11). As Demandt himself has offered convincing counterarguments throughout his book, it is not necessary to evaluate all of these objections in greater detail here. The first objection, however, is so fundamental that it seems worthwhile to briefly dwell on it. Demandt acknowledges that referring to the meaninglessness of the counterfactual principally makes sense. He writes: “Wenn schon ein großer Teil des tatsächlich Geschehenen ohne weitreichende Bedeutung geblieben ist, wie soll dann etwas bedeutend sein, das nicht einmal tatsächlich war? Was nicht stattgefunden hat, hatte auch keine Ursachen und keine Folgen, die es interessant machen könnten“ (Demandt 1986, 11). But later in the book, he seeks to invalidate the objection by emphasizing the important role of counterfactuals as contrast and comparison. In doing so, however, he actually acknowledges the validity of objection, when he writes: “Gewiß bleibt der Einwand richtig, daß das, was nicht stattgefunden hat, auch keine Bedeutung besitzt. Die nicht eingetretenen Möglichkeiten haben selbst keinen Belang, liefern uns aber die notwendige Folie, vor der wir die Bedeutung des wirklich Geschehenen erst erkennen. Irrealität ist ebensowenig ein Argument für Irrelevanz wie Realität kein Argument für Relevanz ist“ (Demandt 1986, 39). While I share Demandt’s conviction that counterfactual scenarios can provide us with a contrast against which the actual meaning of a historical event or condition becomes discernable (we will return to this below), I would like to emphasise that counterfactual alternatives can – under favourable circumstances – have very real consequences as well.

While speaking mainly about causation, Ferguson already hints at this when he writes: “For dice, the past really does not influence the present; all that matters are the equations which govern their motion when thrown. But for human beings the past often does have an influence. To take a simple example (borrowed from game theory): a politician who has shirked military confrontation twice may be emboldened to take up arms the third time he is challenged, precisely because of the memory of those humiliations” (Ferguson 2003, 85-6). While this example – to my understanding – mainly works through causal inference and a subsequent learning process, it nevertheless neatly illustrates how *das Ungeschehene* can, indeed, influence the future course of history – by

detouring via the human protagonist and his/her experiences and perceptions. In Ferguson's example, the politician's judgement of the past will certainly be influenced by his/her thoughts about alternative scenarios. Would a different course of action in the previous challenges have brought a more agreeable outcome? If so, a causal relation has been validated and a lesson learned. But unrealized alternatives that seemed likely to happen to a protagonist can influence decision-taking in a less obvious manner as well.

On previous pages it has already briefly been outlined how counterfactual thinking can invoke an alternative reality that creates a contrast to the factual situation. This contrast can influence how someone feels about a situation. It can, therefore, alter the perception and meaning of a situation. As Orit Tykocinski and Thane Pittman have shown in a series of controlled psychological experiments, counterfactual thoughts can also influence future decisions of the thinkers (beyond the learning processes stimulated by the identification of causal relations, e.g. in the example of the failed exam) (Tykocinski and Pittman 1998). To the thinker, the counterfactual scenario constitutes a possible but not realized alternative against which the actual situation is measured. This process of measurement and comparison can then alter the protagonist's feelings as well as actions. Due to the same mechanisms, contrast inference can be a very valuable tool for the historian as well. As briefly pointed out above in the World War II example, the actions and convictions of historical agents have in many situations been shaped by what they themselves have seen as possible alternative realities. Unfortunately, I am not aware of any historical studies on this particular question on which I could now draw. The following example is, therefore, fictional and somewhat constructed. I hope, however, that it can illustrate how invoking counterfactual contrast can enhance our understanding of the factual course of history in many cases.

Think, for instance, of a group of people unhappy with the *status quo* in a political entity. Their visions of the past and of possible alternatives to reality can become very powerful and influential on their opinions and actions. Their counterfactual thoughts about the past can provide a contrast against which the *status quo* is measured. New, additional meaning is ascribed to the factual situation courtesy of *das Ungeschehene*. In many cases this might not have direct consequences. In other cases, however, the present situation might become unbearable in the light of missed alternatives. The group might be radicalized and stirred into action. In a volatile situation, a counterfactual past might have contributed to creating a group of revolutionaries (or terrorists, depending on one's position in the system). Vice-versa revolutionary groups might – at least for a time – not thrive in a country that has just avoided a particularly disagreeable *downward counterfactual* as the assessment of the current situation is measured against the alternative and generally perceived as not so bad. This is, of course, a simplistic and constructed example and should by no means be accepted at face value in this form. It does, however, illustrate the

connection between counterfactual contrast and historical action: people act on the basis of their perceptions, notions and convictions. They ascribe meaning to events and conditions and are in return influenced by such meanings. These meanings, in turn, depend to a good part on a contrast with possible (maybe even probable) but unrealized alternatives. If the historian wants to go beyond a mere description of past events and seeks to understand convictions and motivations, he/she has to trace the counterfactual thoughts of the protagonists.

For the selection of suitable contrast counterfactuals, Ferguson's rule regarding alternatives "*that contemporaries actually considered*" (Ferguson 2003, 86) can be slightly adapted. Ferguson's contemporaries – from their perspective – thought about possible courses of the future *before* the so-called point of bifurcation. If we want to know how unrealized alternatives change meanings and perceptions, we will mainly be interested in how historical protagonists think about past alternatives and which counterfactual courses seemed likely to them *after* the point of bifurcation. It is important to see that in both cases the alternatives need not seem likely to the historian, but must have appeared realistic and probable to the historical actors under scrutiny. If this can be based on historical evidence, as Ferguson stipulates, the counterfactual enquiry stands on a solid footing.

As illustrated by the two quotes from Demandt's *Ungeschehene Geschichte*, the invocation of contrast by means of counterfactual thinking can have another beneficial effect. While Demandt would reject that unrealized alternatives can develop (or trigger) historical agency, he acknowledges their role in providing contrast (Demandt 1986, 39). Here, it is the alternative that seems possible or relevant to the historian which provides the contrast against which actual developments are measured. Demandt uses value judgements as his example and puts forward that past developments can only be judged in the light of their alternatives. "Ein begrüßenswertes Ereignis ist nur dann erfreulich, wenn es nicht ein noch wünschenswerteres verdrängt hat. Ein bedauerlicher Vorfall ist nur dann schlimm, wenn er nicht das kleinere Übel war und nicht ein größeres Unglück verhütet hat. Zu jedem umstrittenen Faktum müssen wir, wenn wir es werten, die nächstwahrscheinliche Alternative hinzudenken" (Demandt 1986, 25). Demandt mainly refers to moral value judgements, but the same rationale can be applied to, for instance, judgements about economic value as well. In 1964, the future Nobel laureate Robert Fogel published a cliometric study called *Railroads and American Economic Growth: Essays in Econometric History* (Fogel 1964). In the first two parts of the book, Fogel contrasts the real course of history with a counterfactual timeline in which railroads were not available. In comparing the two scenarios, he finds that in many ways the progressive impact of the railroads has been overestimated and that in a world without railroads alternative means of transportation would have led to only a slightly lower performance. Fogel uses an explicit counterfactual as well as clearly defined cliometric methodology and arrives at an insight that made

railroad technology and its economic impact appear in an entirely new light. In a less elaborate form, Alexander Field has attempted a similar study of optical telegraphy in late eighteenth and nineteenth century France (Field 1994). Only the invocation of such contrasting alternatives allows the historian to accurately evaluate the meaning and impact of a historical event, condition or – in these cases – technology.

‘Spielraum’

Historians look back at the past from where they stand in the present. The general course of history is known and always culminates in the present situation – no matter how fleeting such a culmination might be. While the exact mechanisms that brought a particular event or condition about might still be unknown, the result itself is more or less out in the open. Knowing about the outcomes of a historical process, however, changes the way in which historians deal with their subject. Which research questions must be asked? Which factors have been decisive, which ones can be disregarded? All this is decided in the light of a known outcome, a particular course of history. Thinking about counterfactual scenarios provides an alternative course of events that might lead to other questions of which one might not have thought previously or which might not even be relevant regarding to the actual course of history – depending on the degree of plausibility of the counterfactual. In some cases – especially when the alternative is altogether plausible – this practice can open up hitherto neglected paths of enquiry which might indeed lead to unexpected insights. In this respect, counterfactual thinking can create *Spielraum* that re-installs a degree of outcome-openness in historical enquiry.

The creation of *Spielraum* can also have a much more trivial effect, which should, however, not be disregarded here. Thinking about counterfactual scenarios – be they close and plausible or be they downright absurd – can provide welcome distraction from the daily necessities of historical research. As the use of the verb *spielen* suggests, it creates a space to play around and to experiment without the pressure of having to produce results – even without having to adhere to scientific standards. Even if their immediate scientific output might be marginal, such *Spielraum* counterfactuals can bring playfulness back to historical enquiry – a precondition of creativity and innovation that has often gone missing in a working environment that is largely governed by the outcome-orientation of funding bodies and evaluation committees.

Self-awareness

Counterfactual thoughts often hold implicit information about the character and attitude of the thinker. In everyday life, for instance, different people might generate very different counterfactual thoughts (often *upward*, sometimes *downward*) in the same situation. This can provide clues about the notions and

maybe even the personality of the thinker – and in historical research, it can become a helpful tool to handle a researcher’s presumptions and possible biases that are always present in his/her work. Individual views about history and the forces that govern its course can be very different – among historical protagonists as well as among historians. They can range from the belief that everything in history is somehow predetermined to the notion that history is utterly chaotic and contingent. Such ideas about the rationale of human history are often connected with ideological or religious beliefs or the rejection thereof. They are part of our individual *Weltbild* – our image of the world and its inner workings and purpose – and strongly influence our interpretations of past events, causal relations and larger mechanisms. This is true for the casually interested observer of the past as well as for the professionally trained historian. For both, it is not possible (and, indeed, not desirable) to be perfectly neutral in their assessment and interpretation of the past. Their presumptions and their *Weltbild* will always co-shape their interpretations because they provide the very theoretical and moral models to which history is constantly compared. The author will always be present in his/her writing (about) history.

That personal attitudes and beliefs influence the historian’s interpretation of history is not a problem in itself, but it needs to be acknowledged. Students, colleagues or the readers of a book do not need to share the interpretations of a historian, but they should be able to follow their genesis. Therefore, historians should be as explicit as possible about their convictions in order to allow for maximum transparency. This, however, necessitates that the historian himself/herself is aware of his/her implicit attitudes regarding the course of history. Here, developing and fleshing out counterfactual alternatives can be a helpful tool to locate oneself in the wide spectrum between strict determinism and pure contingency. As Niall Ferguson demonstrates, one’s opinion of historical counterfactuals is usually closely related to one’s general view of the laws of history (Ferguson 2003). The well-known controversy between Carr and Isaiah Berlin illustrates this connection. While Berlin fully acknowledged the role of the individual, of agency and accident in history, for Carr, many of the key historical events of the twentieth century – and especially the Russian Revolution – were predetermined and largely inevitable (Carr 1990, Berlin 1997). Accordingly, counterfactual thoughts about an alternative course of history were to him a mere “parlour game” without any practical use as such alternatives could never have happened anyway. Carr’s dismissive opinion of counterfactual history, therefore, already hints at his philosophy of history and helps his readers and students to classify and assess his work.

Even if Carr had not been as dismissive of counterfactuals and could have been persuaded to identify a potential point of bifurcation himself and then develop an alternative timeline, the counterfactuals he would have come up with would have been very insightful as regards his take on history. Chances are that no matter which historical event or person he would have altered or

taken out of the story altogether, in his alternative the current of history would have returned to its predetermined course quite quickly after this small digression. While he unfortunately cannot comment on this anymore, it is reasonable to assume that Carr would have subordinated individual agency and sheer chance (the factors usually manipulated by counterfactual thinking) to bigger historical forces steering towards revolution. In his contribution to this volume, Richard Lebow shows how many social scientists and historians have similar views about the outbreak of World War I. In their views, preventing the assassination of Franz Ferdinand would not have prevented the war as some other provocation or incident would certainly have followed and lit Europe like “dry kindling” (Lebow 2009, 59). Counterfactuals that return to the real timeline after allowing only a small disturbance of the natural course of history hint at a philosophy of history featuring at least some degree of determinism and orderedness.

Believers in contingency and a rather chaotic pattern of history will, on the whole, be less dismissive of counterfactual history than determinists. And they will certainly produce entirely different counterfactual timelines that often differ considerably from reality. As such historians usually ascribe much more importance to individual agency and chance, counterfactuals will generally be more interesting and worthwhile to them. In short, two historians (with different notions about the role of determination and contingency in history) will usually reach very different results and generate different insights, even if they started from the same counterfactual point of bifurcation. No matter how professional and objective they are, their different takes on history and its *primum mobiles* will be mirrored in the course of their counterfactuals. To a certain extent, their *Weltbild* will become discernable – both for the reader as well as for the researcher himself/herself. In many cases, a historian – if asked whether he/she thinks of himself/herself rather as a believer in contingency or determinism – might not even have a ready answer. Possible opinions on this question are as multilayered and entangled as history is itself. And while many (or all) historians will have some inner opinion or feelings about this, many might not have explicitly thought about the problem at all. Here, developing and running through counterfactual scenarios in one’s research field can be extraordinarily enlightening as the nature of one’s counterfactuals tell as a lot about one’s philosophy of history. I believe that thinking counterfactually at least every now and then during one’s own research can significantly contribute to a researcher’s degree of self-awareness and, therefore, in the long run to an increased objectivity of one’s work.

Dangers of Employing Counterfactual Thinking as a Historical Method

Cold war historiography vividly illustrates how the selection of the counterfactual question to be asked generally anticipates the desired answer. (Timothy Garton Ash).

So far, this article has focused on the potential benefits counterfactual thinking can have in scientific analysis. Building on findings of recent psychological research that illustrate the effects and potential usefulness of counterfactual thoughts in everyday life, I have tried to show how similar effects can be invoked in a more scientific context in order to produce new insights. Some of these uses, for instance the testing of causal relations, have been well-known and accepted in many academic disciplines (e.g. philosophy) for quite some time. In the historical sciences, however, counterfactual thinking has received a lot of criticism and has been largely marginalized. The often-cited opinions of E. H. Carr (“parlour game”, “red herring”) and E. P. Thompson (“Geschichtswissenschaft”, “unhistorical shit”) about counterfactual history neatly epitomize how the practice has long been eyed suspiciously within the profession. It was – and to a certain degree still is – a wide-spread belief among historians that history is solely concerned with historical facts, events and processes that actually happened in the past. Manipulating such historical instances removes them from the domain of history. They are of no interest to the historian anymore and can, at best, be used to entertain and distract. Counterfactual history becomes a “parlour game”. In addition, as we have seen above, counterfactual reasoning is particularly hard to reconcile with determinist views on history. In combination, such notions about history and the proper subject of the historian have for quite some time marginalized (explicit) counterfactual thinking in the discipline.

However, while Carr’s and Thompson’s contributions to the field have been invaluable, the discipline has evolved quite quickly ever since and incorporated new ideas popularized by the so-called cultural turns. The Rankean search for “how it actually was” has since been extended. Historians today look at constructions of history, cultural representations or invented traditions just as much as they are interested in the Rankean ideal. Roughly at the same time (but relatively unconnected from the cultural turns) historians such as Alexander Demandt (Demandt 1986) in Germany or Niall Ferguson (Ferguson 2003) in the English-speaking world have published powerful pleas for counterfactual thinking in history. Largely thanks to these individual efforts combined with a general paradigm shift in the historical sciences, counterfactual history has lost much of the stigma of being unscientific that had previously been attached to it. And yet, some uneasiness remains – both on the side of practitioners and critical observers. This uneasiness stems, so I believe, from the various risks that counterfactual thinking brings with it and of which practitioners in the field

must carefully steer clear. Unlike traditional historical enquiry, counterfactual history still has no clearly defined instruments or catalogue of methods. Much more than its “factual” counterpart, it depends on the integrity and self-reflection of the researcher who has to choose adequate points of bifurcation and build his/her alternative timeline on reproducible and reasonable assumptions. While the link between counterfactual history and actual historical sources should never be completely severed, the link is obviously weaker than in historical standard practice. The researcher, therefore, has more room for interpretation and creativity than usual. However, he/she has to be aware of a number of pitfalls arising from this freedom that can easily depreciate or ruin the scientific value of a counterfactual study.

Here, only three potentially detrimental effects associated with counterfactual thinking in historical enquiry shall be discussed briefly: hindsight bias, judgments of blame as well as group- and self-protective biases. In psychological research, these are well-known and have been studied thoroughly for everyday-life situations, but they can appear as unintended by-products of counterfactual thinking in history as well and seriously distort the validity of research outcomes in the field. In the extreme, these biases can – consciously or unconsciously – facilitate and legitimize historical revisionism of dubious ideological provenance and scholarly quality. Therefore, both historians and readers alike should be aware of these effects in order to avoid or respectively detect them in historical writing.

Hindsight bias

Neal Roese explains hindsight bias in an everyday-life context as follows: “After the fact, outcomes are often judged to be more certain and predictable than their prospect was in foresight” (Roese 1997, 142). While it has generally been assumed that counterfactual thinking and its *causal inference effect* can help to cushion hindsight bias, it seems that in many cases the opposite is true (Roese 1997, 142; Roese and Olson 1996). In other words, the *ex post* inference of causation by counterfactual thinking can also invoke a hindsight bias that can make a causal chain seem unavoidable. Roese holds “that there are circumstances in which the counterfactual-hindsight relation is positive; that is, counterfactual thinking may increase rather than decrease the hindsight bias. The key to this argument is that counterfactuals yield causal conclusions that explain why an outcome occurred as it did. And explanation breeds certainty” (Roese 1997, 142). Further down the page he goes on by saying that several experiments found “that manipulations that heightened counterfactual thinking also heightened hindsight certainty, but such effects were mediated by causal inferences. Thus, only when a counterfactual is phrased as a conditional and only when the individual grasps the causal implications of this counterfactual conditional does any increase in hindsight magnitude occur. This evidence

highlights a key facet of the functional argument, that counterfactuals influence causal inferences that may ultimately prove useful, while evoking exaggerated (and perhaps self-serving) certainty regarding the predictability of the past. The latter bias seems a small price to pay for the former benefit” (Roese 1997, 142).

In everyday-life situations, an increased sense of the unavoidability of a particular outcome might indeed be a small price to pay, when causal inference at the same time helps the thinker to learn from the causal chain and its ultimate outcome. Historians, however, are primarily concerned with interpreting historical events in a larger socio-cultural framework and try to make sense of (and give sense to) past events and processes. Here, hindsight bias – no matter if it has been caused or amplified by counterfactual thinking or not – can seriously distort the validity of historical research. Historical events and causal chains might seem much more predictable and unavoidable to the researcher than they actually were at the time – or than they appeared to be to contemporaries. This has implications for the valuation of historical events and actors and their role in history. And it also influences the discussion about the respective roles of contingency and determinism – and, therefore, ultimately about the philosophy of history. It is important to note that hindsight bias originates mainly from projecting *ex post* knowledge about the course of history into the past. It is not a prerogative of counterfactual history and can be found in traditional history as well. Practitioners of counterfactual thinking should, however, be aware that the practice can under circumstances enhance hindsight bias – and to the historian this is certainly not an acceptable price to pay.

Judgments of blame

Similar caution is necessary in order to avoid biases in judgments of blame. Unearthing causal relations and thereby identifying the principal cause of a disagreeable chain of events can lead to blaming someone or something for what has happened. Causal inference through counterfactual thinking, however, can induce irrational judgments of blame that identify abnormal behaviour in a particular situation as the principal cause of events. The classic example revolves around a jogger who is mugged while running a route he rarely takes. “[O]bservers are likely to think that, if he had stuck to his usual route, he would have avoided being mugged. Observers may also note the causal link between [the jogger’s] route choice and his victimization and blame [the jogger] for his misfortune” (Roese 1997, 142). Irrational attributions of blame do not necessarily need counterfactuals but rather thrive on erroneous causal links. And these in turn can be produced by the *causal inference effect* of badly-devised or ill-selected counterfactuals devoid of any analytical quality.

In historical research, blame – in the very sense of the word – should be dealt out sparsely, if at all. While historians naturally do have personal opinions and positions that influence their work and often shine through in their writing,

it is their business to objectively analyse rather than to blame (or applaud) someone for having done something. However, the same mechanisms that provoke biased judgments of blame in everyday life, can distort historical research results as well. Historical agency might be ascribed to someone or something just on basis of his/her/its unusual, unconscious or unintended behaviour. According to chaos theory, a butterfly can cause a whirlwind – and still the butterfly has no agency in the historical-philosophical sense regarding the creation of the whirlwind. Agency depends on the capacity to make choices. While it is a conscious choice to flap one’s wings, the connection between the two events is so contingent that there can be no conscious agency in this respect. The butterfly displays agency as regards flying or moving forward, but not as regards causing a whirlwind. A counterfactual test, however, would reveal a clear causal link and possibly identify agency where there is none. Other factors necessary for the whirlwind to emerge are overlooked. What about the general atmospheric conditions? What about the plane that caused minor turbulences in the air that were later enhanced by the butterfly flap? Without these and many other factors in place, the butterfly effect could not have unfolded. While it is the unusualness of the jogger’s behaviour that provokes judgments of blame, it is the triviality of the butterfly flap that tempts us to ascribe disproportionately high meaning to the actions (the agency) of the butterfly. Both biased ascriptions are the products of causal links inferred by ill-selected counterfactuals. While the *Spielraum* benefit of a counterfactual scenario without butterflies certainly exists, its analytical potential is limited and distracts us from focusing on other causes. In historical research, this can lead to overemphasising trivialities and attributing to them a degree of agency that they often do not have or that can only unfold on the basis of other underlying preconditions.

Group- and self-protective biases

In a recent psychological study on group- and self-protection in counterfactual thoughts, Sean McCrea writes that “[i]ndividuals tend to attribute their own success to internal causes but their failure to external causes, whereas this pattern is not observed for the outcomes of others” (McCrea 2007, 1257). In two unconnected experiments McCrea then goes on to demonstrate that such self-protective thinking can also be seen in the production of *upward counterfactuals* after a disagreeable outcome. In the first experiment, test persons have been asked to list counterfactual thoughts about a negative outcome that had either happened to them or to a dissimilar other person. “Individuals demonstrated a self-protective bias by generating more internal than external disadvantageous counterfactual thoughts for the dissimilar other but not for the self” (McCrea 2007, 1263). In other words, the test persons’ counterfactuals laid the blame for the failure significantly more often on the acting person when this

person was a dissimilar, unknown other than when they themselves were the acting person. The test persons protected themselves in their counterfactuals. The second experiment reproduced this result and extended it to the protection of a group that the test persons strongly identified with. In the counterfactuals listed by the probands, failure of their identification group was significantly less often attributed to internal factors than a similar failure of a group with which they did not identify (McCrea 2007, 1267). This illustrates that the nature of the counterfactual thoughts produced in a particular situation can differ largely depending on the thinker's affections and feelings towards the situation itself but especially towards the acting groups or individuals. In historical research, such a group- or self-protective bias can potentially influence the selection of counterfactuals as well as the nature and course of the alternative timeline. Historians are by no means free of identifying themselves with certain ideologies, social or economic movements, parties or other historical and present groups. This can, of course, lead to unintended and unconscious partiality in their research – no matter if counterfactual methods are employed or not. In counterfactual historical research, however, such partiality might be harder to detect as the actual research is further away from the historical sources and depends even more on the interpretations and views of the researcher. The famous quote from Timothy Garton Ash that opens this section illustrates the potential danger. Practitioners and readers alike should, therefore, be aware of the potential effects of group- and self-protective biases which can subtly influence a researcher's counterfactual choices.

Conclusion

Obwohl nun das physische und moralische Übel nicht notwendig sind, so genügt ihre Möglichkeit auf Grund der ewigen Wahrheiten. Und da diese ungeheure Region der Wahrheiten alle Möglichkeiten umschließt, so muß es unendlich viele mögliche Welten geben, muß das Übel in mehrere von ihnen Eingang finden, und muss die beste von allen Welten es enthalten: hierdurch ist Gott bestimmt worden, das Übel zuzulassen. (Gottfried W. Leibniz (1925, 111)).

One of the most powerful objections against counterfactual reasoning in historical research has been discussed by Alexander Demandt – as *advocatus diaboli* – in *Ungeschehene Geschichte*. Thinking about what might have been in history is inappropriate because it breaks the connection between the past and the present. Demandt writes: “Ändern wir den Kurs der Geschichte an irgendeinem Punkt, dann erreicht sie niemals den Ort, wo sie sich heute befindet. Mit der Folgezeit verwandelt sich auch unsere Gegenwart, unser persönliches Leben. Eine *andere* Geschichte, wie sie auch hätte stattfinden können, wäre jedenfalls nicht mehr *unsere* Geschichte, nämlich die, deren Produkt wir sind. Hier kommen wir aus dem Akzidentiellen rasch ins Existenzielle. [...]

Sobald wir die historischen Voraussetzungen unserer Urteilkategorien im Gedanken beiseitesetzen, heben wir unsere geistige Existenz auf“ (Demandt 1986, 13). This is, indeed, a powerful argument. All our interpretations of the past – all our attempts to inflict meaning on past events and processes – can only be made from our current perspective. Our own history and experiences determine the questions we ask and the topics we select. Or as Rosalind Williams, adapting a metaphor by Peter Galison, states: “History itself is a trading zone between past and present. History establishes communication between storytellers and actors, between what happened and what is remembered, between the dead and the living” (Williams 2005, 435). Counterfactually manipulating history, therefore, changes the past as much as it changes the present. An interpretation of the past from our current viewpoint becomes impossible as the altered past requires a different viewpoint – which we do not and can never know let alone take.

Does this well-delivered objection negate all that has been brought forward to support counterfactual thinking as a tool for historical enquiry? What can we hold against this argument, when it invalidates the very perspective of counterfactual interpretation? I believe that there is only one thing that we can bring up in defense: namely that the practice should be taken seriously, while at the same time not taking itself too seriously. Counterfactual history should not claim to interpret a counterfactual past along the same rules and guidelines that apply to traditional history. Practitioners must acknowledge that they operate on an incomplete knowledge basis that is even more fragmentary and based on assumption and extrapolation than traditional history already is. We must embrace the fact that our alternate timelines lack a suitable viewpoint for interpretation – and that, therefore, the results of counterfactual reasoning in historical enquiry will always be complementary to other insights. Counterfactual history can offer valuable contrasts and a laboratory to test hypotheses and causal links, but its findings will only make sense in the light and within the framework of traditional historical research.

Because it would be utterly disappointing to finish an article in support of counterfactual thinking with such a comment on the limitations and complementary character of the practice, let me end by reminding the reader of Niall Ferguson’s notion that historical counterfactuals should primarily be concerned with alternatives that seemed possible and plausible at the time and were actually contemplated by contemporaries (Ferguson 2003, 86). While separating the past and the present like any other alternative timeline, such counterfactuals can complement traditional historical enquiry by bringing the perspective of the contemporaries back to life – with all its possibilities, insecurities and open choices. And in doing so, counterfactual history responds to the demand that one of its staunchest critics has voiced in his most influential work – E. P. Thompson in *The Making of the English Working Class*. In the preface,

Thompson states that he is aware of “writing against the weight of prevailing orthodoxies” in his book. And he goes on to say that, among others,

[t]here is the ‘Pilgrim’s Progress’ orthodoxy, in which the period is ransacked for forerunners – pioneers of the Welfare State, progenitors of a Socialist Commonwealth or (more recently) early exemplars of rational industrial relations. [...] My quarrel with [this orthodoxy] is that it reads history in the light of subsequent preoccupations, and not as in fact it occurred. Only the successful (in the sense of those whose aspirations anticipated subsequent evolution) are remembered. The blind alleys, the lost causes, and the losers themselves are forgotten. [...] I am seeking to rescue the poor stockinger, the Luddite cropper, the ‘obsolete’ handloom weaver, the ‘utopian’ artisan, and even the deluded follower of Joanna Southcott, from the enormous condescension of posterity. Their crafts and traditions may have been dying. Their hostility to the new industrialism may have been backward-looking. Their communitarian ideals may have been fantasies. Their insurrectionary conspiracies may have been foolhardy. But they lived through these times of acute social disturbance, and we did not. Their aspirations were valid in terms of their own experience, and if they were casualties of history, they remain, condemned in their own lives, as casualties. [...] Our only criterion of judgment should not be whether or not a man’s actions are justified in the light of subsequent evolution. After all, we are not at the end of social evolution ourselves” (Thompson 1972, 12-3).

References

- Austin, John L. 1970. Ifs and Cans. In *Philosophical Papers*, ed. James O. Urmson and Geoffrey J. Warnock, 205-232. Oxford: Clarendon Press.
- Berlin, Isaiah. 1997. Historical Inevitability. In *The Proper Study of Mankind, An Anthology of Essays*, ed. Isaiah Berlin, 119-190. London: Chatto and Windus.
- Carr, Edward H. 1990. *What is History?* Harmondsworth, Middlesex: Penguin.
- Demandt, Alexander. 1986. *Ungeschehene Geschichte. Ein Traktat über die Frage: Was wäre geschehen, wenn...?* Göttingen: Vandenhoeck & Ruprecht.
- Ferguson, Niall. 2003. Virtual History: Towards a ‘chaotic’ theory of the past. In *Virtual History. Alternatives and Counterfactuals*, ed. Niall Ferguson, 1-90. Basingstoke, Oxford: Pan Macmillan.
- Field, Alexander J. 1994. French Optical Telegraphy, 1793-1855: Hardware, Software, Administration. *Technology and Culture* 35: 315-47.
- Fogel, Robert W. 1964. *Railroads and American Economic Growth: Essays in Econometric History*. Baltimore: Johns Hopkins Press.
- Goodman, Nelson. 1947. The Problem of Counterfactual Conditionals. *The Journal of Philosophy* 44: 113-128.
- Halpern, Joseph Y., and Judea Pearl. 2005. Causes and Explanations: A Structural-Model Approach. Part I: Causes. *British Journal for the Philosophy of Science* 56: 843-87.
- Lebow, Richard N. 2009. Counterfactuals, history and fiction. *Historical Social Research/ Historische Sozialforschung* 34, *Special Issue on Counterfactual Thinking*, ed. Roland Wenzlhuemer, 34.2, 57-73.

- Leibniz, Gottfried W. 1925. *Die Theodicee. Neu übersetzt und mit Einleitung Anmerkungen und Register versehen von Artur Buchenau*. Leipzig: Meiner.
- Lewis, David. 1973. Causation. *The Journal of Philosophy* 70: 556-67.
- Markman, Keith D., Igor Gavanski, Stephen J. Sherman, and Matthew N. McMullen. 1993. The mental simulation of better and worse possible worlds. *Journal of Experimental Social Psychology* 29: 87-109.
- McCrea, Sean M. 2007. Counterfactual thinking following negative outcomes: Evidence for group and self-protective biases. *European Journal of Social Psychology* 37: 1256-1271.
- Morello, Ruth. 2002. Livy's Alexander digression (IX. 17-19): counterfactuals and apologetics. *Journal of Roman Studies* 92: 62-85.
- Roese, Neal J., and James M. Olson. 1996. Counterfactuals, causal attributions, and the hindsight bias: A conceptual integration *Journal of Experimental Social Psychology* 32: 197-227.
- Roese, Neal. 1997. Counterfactual Thinking. *Psychological Bulletin* 121: 133-48.
- Thompson, Edward P. 1972. *The Making of the English Working Class*. Harmondsworth, Ringwood: Penguin Books.
- Tycocinski, Orit E., and Thane S. Pittman. 1998. The Consequences of Doing Nothing: Inaction Inertia as Avoidance of Anticipated Counterfactual Regret. *Journal of Personality and Social Psychology* 75: 607-16.
- Williams, Rosalind. 2005. Afterword: an historian's view on the network society. In *The Network Society: A Cross-cultural Perspective*, ed. Manuel Castells, 432-448. Cheltenham: Edward Elgar.