AUTHOR(S)
Virmajoki, V.

TITLE
On the Function and Nature of Historical Counterfactuals. Clarifying Confusions

YEAR
2024

DOI
10.1163/18722636-12341519

CITATION
Virmajoki, V. (2024). On the Function and Nature of Historical Counterfactuals. Clarifying Confusions. *Journal of the Philosophy of History*, 1–25. https://doi.org/10.1163/18722636-12341519

VERSION
Accepted Manuscript

On the Function and Nature of Historical Counterfactuals. Clarifying Confusions

Abstract

In this article, I analyze historical counterfactuals. Historical counterfactuals are conditional statements in which the antecedent refers to some change in the past. We ask what would have happened, had that change occurred. I discuss the nature of such counterfactuals. I then identify important functions that historical counterfactuals have. I point out that they are at the heart of explanations and, therefore, reveal issues related to contingency and actual history. I then discuss counterfactual reasoning in historiography. I argue that the problem of suitable antecedent conditions has been exaggerated, and more serious issues concern the tracking of counterfactual scenarios. Throughout the paper, I argue that the interventionist way of thinking about historical counterfactuals clarifies both historical explanations and the nature of historical counterfactuals and should be adopted as the standard. I conclude by noting that historical counterfactuals may not fundamentally differ from more familiar forms of historiography.

Counterfactuals – Contingency – Explanation – Generalization – Narratives – Minimal Rewrite Rule

1. Introduction

"Had Eddington lost his faith in humanity, the (amount of) gravitational deflection would not have been observed"; "Had the Snowball Earth been different, the Cambrian explosion would not have occurred". These statements are about events that did not happen. The antecedents are counter to the facts – it is *counterfactual*. The whole statements are *counterfactual conditionals* that describe how things would have gone, had something gone differently. For simplicity, we will use the term "counterfactual" to refer to counterfactual conditionals. *Historical counterfactuals*, then, are counterfactuals that are relevant in historiography and

historical sciences, i.e., in the fields that are focused on particular occurrences (Glennan 2010; Gangl 2021).

In historiography, the most interesting counterfactuals concern token events. They are about one particular event (or condition or pattern) like the observation of the gravitational deflection. In historiography, we are interested in a particular event being different and the consequences for the history that this difference makes. Surely, this may enable us to achieve insights on many other events as well or even on event-types (how the type of change in question works in general), but we shall focus on counterfactuals concerning event-tokens as they are the ones that have gained the most attention in the philosophy of historiography.

In this article, we discuss the functions that counterfactuals have in historiography and analyze the principles that govern counterfactual reasoning. The questions are (i) how counterfactuals function in historiography and (ii) how we can reason about counterfactuals in a controlled manner. As Tucker puts it, "Philosophers and historians debate not only the correct analysis of historiographic counterfactuals and their possible utilities for historiography and its philosophy, but whether they can amount to more than speculations and have any use at all".[1] We can already note a bit of a dilemma with respect to historical counterfactuals: On the one hand, they have several central functions ranging from establishing contingency to grounding explanations. We cannot live without them. On the other hand, the epistemology and even the motivation for historical counterfactuals have been constantly questioned. Carr famously called them "parlour games".[2] More recently, Evans argued that "counterfactual speculations frequently, perhaps even generally, tread on thin evidential ice, too often selecting the conditions of their starting point with insufficient care,

[1] A. Tucker, "Historiographic Counterfactuals and the Philosophy of Historiography. An Introduction," *Journal of the Philosophy of History* 10, no. 3 (2016): 333-348.
[2] E.H. Carr, *What Is History?* (New York: Vintage, 1961), 97.

and failing to distinguish between different levels of causation".[3] In what follows, I discuss the functions of historical counterfactuals and, on the basis of this discussion, I discuss the epistemology and motivations for historical counterfactuals.

The overarching theme in the paper is the argument that (i) because historical explanations require counterfactual thinking, (ii) because this explanatory function can be defined through interventionist account of explanation,[4] and (iii) because interventionism can also clarify many issues related to counterfactual reasoning beyond explanation, (iv) interventionist understanding of counterfactuals should become the standard starting point. This means that I go beyond the common suggestion that counterfactuals and explanation are connected. I add that there is a specific theory of explanation whose conceptual tools can be used in counterfactual reasoning throughout its functions.

I proceed as follows. In Section 2, I briefly introduce historical counterfactuals by providing two illustrations, one from historical sciences and one from historiography. I use an example from historical sciences in this paper in order to (i) have the ability to introduce counterfactuals in the context that is not as loaded with our historiographical reflexes as human history, and (ii) have a case against which to compare historiographical counterfactuals when needed. In Section 3, I discuss the different functions that counterfactuals are considered to have in historiography. I also discuss what types of counterfactuals are needed for the central functions to work. In Section 4, I discuss the problems associated with counterfactuals. I point out that the philosophical literature is misleading in many places with regard to the issue. Most analyses have focused on the question of how the antecedent of a counterfactual should be understood. There has been less

[3] R. Evans, *Altered Pasts: Counterfactuals in History* (Waltham, Mass: Brandeis University Press, 2013), 148.
[4] J. Woodward, *Making Things Happen: A Theory of Causal Explanations* (Oxford: Oxford University Press, 2003).

debate on how to track the development from the antecedent to the consequent of the conditional. I point out that the latter is the interesting question, not the former. Throughout the paper, I argue that interventionist account can be used to clarify the philosophical literature on historical counterfactuals.

2. Historical Counterfactuals

As we already noted, counterfactuals are conditionals where the antecedent is false. They tell us what (supposedly) would have happened in the case that the antecedent happened. The important questions are (i) how, exactly, the antecedent is supposed to have happened in the scenario, and (ii) how we connect the antecedent to the consequent. While we return to these questions in later sections, it is beneficial if we understand the fundamental issues related to these questions already.

First, we can sometimes build scenarios in which the antecedent happens in a way that alters the outcome in a misleading way. For example, in most contexts, the claim "Had the barometer reading dropped, a storm would have occurred" is misleading. If the barometer reading is supposed to drop because the atmospheric pressure drops, then the counterfactual is true but we find it misleading. There is something wrong with the type of change that changes the antecedent in the scenario. Moreover, in some cases, changing the antecedent might be meaningless and the whole counterfactual requires a novel interpretation. For example, if we say "Had I been a dog, I would have eaten the same food all the time", there is the question of what would it even mean to change me into a dog. To sum up, we can conceive all sorts of possibilities, but it is difficult to tell which of these are relevant for counterfactual considerations.

The second question is a difficult one because we cannot have, by definition, any direct evidence about counterfactual scenarios, i.e., how things would have developed, had

something been different. This problem has often led to a dismissive attitude towards historical counterfactuals and the idea that they are mere "parlour games". The fact that philosophical literature on counterfactuals has not been all that interested in the epistemic details of how counterfactuals in any real-life contexts such as history can be traced does not help the issue. For example, David Lewis suggested that once the antecedent changes in the right way, the scenario develops in accordance with the laws of nature (1979, 463). This hardly tells anything about how counterfactuals can be tracked in historiography. However, given the important functions that historical counterfactuals have, we should resist the dismissive attitude. I return to this below.

Now, to put some flesh on the bones, consider two illustrations.

First, consider the event in 1919 when Arthur Eddington observed the amount of gravitational deflection. The deflection matched the prediction of Albert Einstein's theory of relativity. What does the history of this event look like? Long story short: Eddington was born into a Quaker family. Eddington's Quakerism affected his will to unite the international scientific community during and after the Great War (Stanley 2003; 2019). On a general level, the war cut all connections between Allied and Central powers, including those of science. It led to deep and long-lasting suspicion among the scientists. In Britain, there were widespread plans to exclude Germany from science even after the war (Stanley 2019, 176-181). In 1905, Einstein published his special theory of relativity. He then extended his work to gravity and started to develop the general theory of relativity. A consequence of his theoretical reasoning was that gravity should bend light to a specific degree (gravitational deflection). In 1915, Eddington wrote about gravitation and Einstein's prediction. Eddington became interested in the theory but also in the fact that it was formulated by a pacifist in Germany. (Stanley 2019, 168-270). He wished to use the theory to bring British and German science closer again. (Stanley 2003; 2019.) An opportunity to make Einstein's theory famous was a solar eclipse in

1919 during which the deflection could be observed by taking photographs of stars around the sun. Eddington had to convince the British scientist about the importance of the theory, and, finally, was able to execute an expedition to take the photographs. (Stanley 2003; 2019.) After analyzing the data, Eddington declared that the deflection is confirmed.

We can ask many questions about counterfactual developments of the scenario. For example, we could ask, what would have happened, had Eddington not been a Quaker. Would the observation of the deflection have been made in 1919? Prima facie, it seems that the observation would not have been made, had Eddington not been a Quaker. His religious background gave him a unique incentive in British science to focus on Einstein's theory. However, it seems a bit unclear what we mean by the antecedent. Do we assume that Eddington had some other religious beliefs or that he had lost his faith in humanity? These two scenarios lead to vastly different outcomes. Moreover, how can we track the scenario where Eddington did not have the religious beliefs he had? How do we know that no one else would have made the observation in 1919? We can see that many interesting issues arise with respect to counterfactuals.

Secondly, consider the fact that there have been periods (hundreds of millions of years ago) when the Earth was covered with ice, the Earth being a "Snowball". It is interesting that "the latest snowball event occurred relatively soon before the Cambrian explosion, the (geologically) sudden rise of complex metazoans [multicellular life]" (Currie 2018, 55). "It is generally accepted that these events are linked: The ancestors of Cambrian fauna must have survived Snowball Earth" (ibid. 153). However, there is an even stronger connection: "For the radiation to occur, pockets of life must have been isolated in order to diverge both phylogenetically and developmentally, without evolving complex, novel traits" (ibid. 154). This means that "Aspects of the Neoproterozoic glaciation and the Cambrian explosion, then,

depend on each other. … *Had the glaciation been different, so too would the explosion have been different*." (ibid. 154, emphasis added).

Again, we see that two events are connected through counterfactual dependence. As in the Eddington case, there is a story that can be told on why the dependence holds. For example, the glaciation isolated pockets of life thus allowing divergence. We can see that, while the timescales of the two scenarios differ enormously, there is a similar underlying logic in how they work. However, unlike in the Eddington case, it does not seem unclear what we mean by the counterfactual alternatives to the glaciation. We are discussing changes in how frozen the Earth was. What counts as a relevant amount of change can be debated, but different readings of the antecedent do not pull to wildly different directions.

The two examples above illustrate how historical counterfactuals arise from our knowledge about particular events. In both examples, the reasoning behind counterfactuals appear to be rather controlled – they do not involve wildly speculative reasoning. Moreover, while there might be some obscurity in the antecedents, the problems of understanding the starting points of the relevant scenarios do not seem overwhelming. We will contrast these observations with philosophical problematization in Section 3. Before that, we will look at what counterfactuals can be used for.

3. The Functions of Historical Counterfactuals

It is completely natural to think that historical counterfactuals are important to be considered for their own sake. We might simply be interested in what would have happened, had the Snowball Earth not occurred or had Eddington not made to observation thus making Einstein a scientific star. In general, we might wonder about questions that are intimately connected to counterfactuals such as "Could science have been different from what it actually is" (see Soler et al. 2015) and believe that such counterfactuals give us insights on the domains of

life, such as science. To me, this seems like a legitimate function for counterfactuals. However, given the widespread skepticism toward historical counterfactuals, this answer does not lead us very far. If there is no legitimate function besides intellectual games of curiosity, it seems difficult to motivate the analysis of epistemically sound and controlled counterfactual reasoning.

Historical counterfactuals do serve important functions. We can identify two kinds of functions, *practical* and *epistemological*. The *practical* function of historical counterfactuals relates to their ability to provide insights that are significant to our self-understanding or even moral issues. For example, the fact that, had the glaciation occurred differently, the evolution could have occurred differently, gives an insight into the place of human beings in nature. Moreover, Einstein has been a scientific icon for over a century now – what if Eddington never raised Einstein to this position? What would our scientific imagination look like? Interestingly, Rosenfeld (2016) has identified several categories of counterfactuals, one of which is *emotive*. As Rosenfeld points out, counterfactuals can give us insights into issues such as missed opportunities that may introduce the feeling of regret. However, while the practical function of counterfactuals is important for us human beings and motivates the need to tell how such insights are achieved in counterfactual reasoning, it does not quite tell us why historical counterfactuals are important in research settings.

The *epistemological* function relates to their ability to provide insights that are relevant to our historiographical understanding and knowledge. These are the functions that we focus on in this paper. I will identify three main functions: (i) providing insights on the actual course of history, (ii) providing insights on contingency, (iii) providing explanatory understanding. I will point out that historical counterfactuals are able to satisfy functions (i) and (ii) through function (iii).

First, it can be argued that counterfactuals are useful enhancements for historical reasoning about the *actual* course of events. In general, one could say that counterfactual thinking is a good exercise to open up historiographical imagination and leads to novel insights. However, this does not answer the questions of why there should be any epistemic credibility in the scenarios thus constructed (we can open up our imagination by using all sorts of intellectual games) and how, exactly, opening up imagination serves sound historiography.

There have been more detailed analyses of the relationship between counterfactual considerations and the study of the actual history. For example, Nolan argues that counterfactual thinking enables the historian to recognize how the history might or might not have gone and, thereby, identify those paths that best fit the evidence (446-447). There is also a tendency to claim that counterfactual histories enable historians to identify *actual* causal chains by identifying how changes in the events in those chains would have changed the outcome. In these arguments, counterfactual reasoning is considered merely as one method (among others) that can be used to establish actual causal chains. For example, Tucker argues that "there are probably several methods historians use to justify causal explanations of descriptions of events. The counterfactual method is one such method." (2016, 341). Similarly, Woolf argues that counterfactual scenarios can be a heuristic tool to establish actual causal sequences: "Counterfactual is something like a scaffolding surrounding the building or rebuilding of an edifice, rather than the edifice itself. Like the scaffolding, it should disappear when the building is complete." (2016, 426-427).

The problem is that, as we will see below, counterfactuals are much more intimately connected to causal explanations in historiography than the perspective above suggests. It is impossible to have an explicit causal explanation in historiography that cites only an actual sequence of history. There needs to be some information associated with causal sequences that separate them from non-causal ones. Information about counterfactual dependencies is a

strong candidate as the right add-on to the actual sequence. As Bulhof puts it, "Counterfactuals, causes, and explanations are three sides of the same strange three-sided coin; you cannot have one without the other two" (1999, 147). This is good news for counterfactuals: Given that even those who remain skeptical (like Evans 2013; 2016) about historical counterfactuals still allow there to be causal explanations in historiography, we can see that we cannot escape from historical counterfactuals. If one thinks that actual causal sequences can be established to provide explanations, one already believes in the possibility of historical counterfactuals. I return to this soon and I then also explain how counterfactuals can matter for the reconstruction of the actual history.

Secondly, it has been suggested many times that counterfactuals can reveal contingency in history. In Ben-Menahem's (1997; 2016) suggestion, historical necessity and contingency should not be mapped one-to-one with causality and randomness. Rather, both notions are causal notions and provide information about the underlying causal structure. They are defined in terms of degrees of sensitivity to initial conditions and intervening factors. The idea is that the more sensitive the outcome is to the changes in the initial conditions and intervening factors, the more contingent it is. For example, Eddington's observation of the deflection was a rather contingent event: had the weather been worse on the day of the eclipse, the outcome would not have been made. There was sensitivity to the initial conditions (bad weather in this case). In a similar spirit, Virmajoki (2018) suggests that the contingency of a historical outcome can be established by constructing a counterfactual scenario that would have led to a different outcome and estimating how plausible that scenario was. What connects counterfactuals and contingency in these types of analysis is the idea that contingency is revealed by estimating how easily things *could have been different*.

One might still wonder why contingency would matter for historiographical methodology. Is there really a need to discuss how contingent the history was? There are two possible

answers. First, it could be said that understanding contingency simply is something that historiography provides and we appreciate. Historiography provides us with perspectives on the world and part of this is due to its ability to identify degrees of contingency. This type of view has been taken in historiography of science. For example, Kohler and Olesko argue that:

"This understanding [that science is a product of the society that creates and harbors it] was achieved not by abstract reasoning [--] but empirically, by detailed study of local sites of knowledge making. These showed concretely how scientific findings were the products of particular local situations and communal practices with all their historical and social contingencies." (2012, 3.)[5]

Secondly, and more seriously, it is important to notice that contingencies are often seen as an obstacle to the reconstruction of historical causal chains. Given that historiography does not have simple stories that are guided by law-like associations, telling the history is difficult. In historiography – or so it seems – we need to take into account all sorts of minor occurrences that might have had a serious effect on the outcome. Ironically, the opponents of historical counterfactuals have argued that, because of the contingencies, counterfactual scenarios are untrackable. For example, Evans argues that "some counterfactual histories rest on an inadequate understanding of historical constraints, ignore *possible contingencies* that might have disrupted a counterfactual timeline" (2016, 462, emphasis added). Given that contingencies are events that could plausibly occur and affect what happens, the construction of scenarios is difficult. However, this rests on the assumption that contingencies matter in the history. And if they matter in history, we surely need to be able to keep an eye on them, and this is possible only through counterfactual thinking.

[5] See also Daston 2009 making a similar argument.

Finally, counterfactuals are central to causal explanations in history. That there is some connection between causal explanation and historical counterfactuals is almost universally agreed (see Weber 1949; Bulhof 1999; Maar 2014, 2016; Ben-Menahem 2016; Sunstein 2016; Bunzl 2004; De Mey & Weber 2003; Lebow 2000; Glennan 2010; Virmajoki 2022). However, as we have seen, some philosophers have questioned whether counterfactuals are essential to explanations in historiography. Next, we see that they are.

In order to approach the issue, we can first notice that some have associated historiographical explanations with contingency. It has been argued that (causal) narrative explanations are especially good at exhibiting contingency. Narrative explanations explain particular events via causal sequences concluding with the *explanandum* (Currie 2014, 1165). Beatty has suggested that "narratives are especially good at representing contingency and accounting for contingent outcomes" (2016, 34; see also Beatty and Carrera 2011). A narrative provides understanding about turning points in a sequence of events leading to the outcome (Gallie 1964; Beatty 2016).

But what makes narrative explanations *explanations,* i.e., how are things connected to each other in an explanation? Whatever the answer, it must be able to capture the central role of contingencies in narrative explanations. The answer must show how turning points are tracked.

There are two relatively recent candidate accounts of explanation that are highly relevant to the issue at hand. One is the so-called mechanistic account of explanation. According to this account, historiographical explanations describe *ephemeral mechanisms* that produce the historical outcome that we wish to explain. An ephemeral mechanism is a short-lived and non-stable collection of interacting parts whose interactions can be characterized by direct, invariant, change-relating generalizations (Glennan 2010, 260). What is important for us is

that the mechanistic account of explanation requires that the change-relating generalizations used in explanations are "counterfactual supporting" (Glennan 2010, 257). In other words, if we cannot rely on counterfactual considerations, we cannot construct historiographical explanations. On the other hand, once we have the counterfactual supporting generalizations, we can track turning-points in narratives.

The other candidate is a *counterfactual account of explanation*. As the name promises, counterfactuals are central to explanations in this account. Counterfactual accounts of explanations have roots deep in the history (e.g., Weber 1949), but one of the most popular accounts today is the so-called interventionist account. This account was developed by Woodward (2003) and applied to historiography by Virmajoki (2022). In this account of explanation, "the underlying or unifying idea in the notion of causal explanation is the idea that an explanation must answer what-if-things-had-been-different questions, or exhibit information about a pattern of dependency" (Woodward 2003, 201). To put it simply, explanations answer questions of the form "Why X rather than Y?" by pointing out factors Z and W such that "had W rather than Z been the case, Y rather than X would have been the case." Explanations provide information about counterfactual dependencies between *explanans* and *explanandum* and are contrastive in nature. An explanation "must enable us to see what sort of difference it would have made for the explanandum if the factors cited in the explanans had been different in various possible ways" (Woodward 2003, 11).

Given the interventionist account of explanation (and counterfactual accounts in general), counterfactuals are at the core of historiographical explanations. Counterfactual considerations are not a mere tool to establish actual causal sequences that explain. In fact, it does not even make sense to say that we could explain through actual causal sequences. A description of a sequence does not include information about patterns of dependencies. Only if we grasp these patterns, we have achieved understanding. We need counterfactual

scenarios in order to have historical explanations; having knowledge of counterfactual scenarios (patterns of dependence) is identical to having an explanation.

This means that those who believe that historiography can provide causal or explanatory insights must accept that historical counterfactuals make sense and can be tracked.

Now, let's work back to where we started from. Given the central role of counterfactuals in historiographical explanations, we can make sense of how we can gain insights into contingencies and the actual history through counterfactuals.

First, given that contingency is related to the sensitivity of the outcome on other factors, contingency can be assessed on the basis of explanations. Given that an explanation provides us with information about counterfactual situations where the outcome would have been different, we can assess the contingency by asking how plausible it is that these situations occurred. Narrative explanations are especially good at representing contingency and accounting for contingent outcomes because explanations in general are suitable for this task. If we understand narratives as tracking counterfactual dependencies between the items in the causal sequence and if contingency is defined in terms of dependencies, then narratives quite naturally are good at representing contingency.

Secondly, explanations are not easily constructed. It is here that counterfactuals have their merits in providing insights into the actual history. First, counterfactual scenarios necessarily require some assumptions about how the world works (see the next section). We need some generalizations or principles in order to track counterfactual scenarios. How well we can track a scenario serves an indicator for what we know and assume about the actual history. For example, if we track the scenario "Had Eddington lost his faith in humanity…" and conclude that he would still have made the expedition, we are relying on the idea that only epistemic values guide scientists. This type of reasoning tells us what generalizations we rely

on when interpreting the history and provides us a way to critically assess the generalizations. Similarly, our inability to track counterfactual scenarios may set new questions and guide the search for evidence. For example, even if we cannot track, based on our current knowledge, the claim "had someone else made the observation the deflection, the community would have remained more cautious towards the theory," making the gaps in our reasoning explicit suggests what generalizations are required to complete it. This limits the types of considerations that are required in the assessment of the explanatory claim and, in best cases, suggests how new evidence could be relevant by telling what evidence would be relevant for the justification of the generalizations.

We have seen that counterfactuals are central and useful in historiography. They are essential for historical explanations and historical contingency. Moreover, the construction of counterfactual-based explanations is useful for the study and interpretation of the actual history. While this latter function could perhaps be achieved by other means as well, as Maar (2014) argues, it is difficult to see why we should not be omnivores (to use Currie's [2018] phrase) when it comes to ways of gaining historiographical insights.

Now it is time to turn to the epistemological issues surrounding counterfactuals. The discussion already suggests that "the difference between counterfactual and 'factual' history may … be marginal", as Lebow (2000, 553) puts it. In the next sections, we elaborate this idea and discuss how counterfactual should be thought of.

4. How to Understand Counterfactual Antecedents?

There are two main questions concerning counterfactual reasoning in historiography. First, how should we understand the change required by the antecedent? This question has received a lot of attention in the philosophy of historical counterfactuals. Secondly, how can we track

what would have happened after the antecedent? This question has received less attention but is more interesting. Yet, we shall proceed by discussing the former question first.

Consider the counterfactual "Had the barometer reading dropped, a storm would have occurred". It seems that the truth of this claim depends on how the barometer reading is supposed to have dropped. If it dropped because the atmospheric pressure dropped, then a storm would have occurred. On the other hand, if it dropped because I adjusted the needle with my finger, a storm would not have occurred. How should we understand the change in the antecedent – the *transition period* as some name it? (Greene 2021.)

In his canonical analysis, stemming from philosophical needs, David Lewis suggested that the antecedent is supposed to be changed by a miracle, the smallest change required for the antecedent to occur (1973). There are technical problems in the suggestion (see Woodward 2003, 133–145) but the philosophy of historiography has focused more on a motivational problem. It has been argued that historical counterfactuals need to be understood in terms of changes that were historically plausible or coherent. In Elster's terms, the antecedent "must be capable of insertion into the real past" (1978, 184).

This line of thought has been common in the literature (see Greene 2021). The thought is expressed in the so-called *minimal rewrite rule* which says that we should avoid far-fetched counterfactuals in historiography and consider only those counterfactual situations that could have happened, given the actual history (see Tetlock and Belkin 1996, 23-25). This rule has two components. It says that (i) the antecedents should be plausible, given what we know about history, and (ii) the connecting principles and steps that we follow to consequent must be in accordance with the actual history. We will discuss (i) now, but we should note that (ii) is related to Section 5 where we discuss the trackability of counterfactuals.

While this rule is no less ambiguous than Lewis's suggestion, it is supposed to respect historiographical counterfactual thinking more. "When implementing a counterfactual antecedent, the historian thus asks what conditions would have to be present in order for the antecedent to follow from these conditions, and whether these conditions were likely." (Reiss 2009, 719). There have been several attempts to make the rule more precise by identifying some *categories* of events whose occurrence supposedly are, in general, likely or plausible.

One of the suggestions has been that we should focus on decisions and scenarios that were considered by the people involved in the historical situations. "We should consider as plausible or probable only those alternatives which we can show on the basis of contemporary evidence that contemporaries actually considered" (Ferguson 1999, 86). In a similar spirit, Tetlock and Belkin suggest that "investigators might agree to constrain counterfactual speculation in a host of more specific ways: by considering as antecedents only those policy options that participants themselves considered" (1996, 23).

However, such attempts to definitively categorize events as likely or plausible are not helpful. For example, consider the idea that we should only consider historical paths that were contemplated by people at the time. This is not always an accurate measure of plausibility. After all, what people considered plausible in the past does not necessarily match up with what was really possible. At the end of the 19th century, fundamental changes in physics were not widely considered plausible. However, we know that the fundamental changes were very much plausible, as they occurred.

An even more serious problem is that these types of considerations confuse the original counterfactual "Had C been the case, E would have been the case" and the counterfactual "Had D been the case, E would have been the case", where D is the most plausible conditions where C was the case. However, given that we are interested in the dependence between C

and E, the dependencies between D and E are irrelevant. And if we assess the dependencies between D and E, we are not drawing conclusions about dependencies between C and E. These issues are logically distant.

The underlying worry behind the suggestions like ones above is that we need to somehow clarify how the antecedent has taken place. The worries have been taken too far. There are perfectly legitimate ways to define how the antecedent was changed that do not require any general insights on what is possible in history and what is not. We can use the notion of a *surgical intervention* from Woodward to define the changes in antecedents. When we make an intervention to change the actual C* to counterfactual C, we change the system in such a way that only C* is changed to C. The intervention does not have an independent effect on other relevant factors (Woodward 2003, 14; see p. 98 for details). This means that when we consider a counterfactual, we consider a case where the antecedent occurs because of this type of stipulated change. We are not asking how the change could "really" have happened. We are interested in what follows from the change from C* to C. This is what our counterfactual "Had C been the case…" is about. If we ask, what would have happened, had Eddington not made the observation, we simply follow through a scenario where the observation is erased by intervention – for example, a good candidate is to assume that the clouds made the starts invisible. We are not asking any surreal questions about changes in wider order of things that led to Eddington's failure in some imaginative way.

Again, we need to be alert here. Surely, sometimes our question about C is a disguised question about D. For example, one might wonder what would have happened, had Eddington not made the observation because the scientific community was taken as a hostage by aliens. In this case, out question is "what would have happened if D rather than (actual) D* was the case" and we answer by considering an intervention. The fact that we may

confuse different questions together hardly means that we cannot have clarity when answering the questions themselves.

Of course, there might be counterfactuals that do not make sense (see Sunstein 2016, 347). In such cases, we do not conceive any intervention that could inject the antecedent into the history. For example, it makes no sense to assume that an intervention could make me a dog. However, there is no reason why such counterfactuals should provide reasons to doubt counterfactuals in general. "Had the Snowball Earth been different, the Cambrian explosion would have been different" is a perfectly legitimate counterfactual while "Had Bill Clinton been a dry match, he would have ignited when struck" is not. I see no reason to assume that most important historical counterfactuals fall into the nonsensical category.

Understanding counterfactuals in the interventionist way is useful also because, as we saw in the previous section, counterfactuals are essential to explanation. C explains E only if E depends causally on C. Not just any counterfactual change is good enough for this purpose. If the barometer reading drops because the atmospheric pressure drops, a storm would occur. The barometer reading still does not explain the storm. The change in atmospheric pressure does not change the reading in a correct way. We can say that it is not an intervention, as it has an independent effect on E that does not go through C. This means that, because we need historical counterfactuals to explain, we need the interventionist interpretation of the "transition period" anyway to construct the relevant counterfactuals. It follows that the interventionist way of thinking is the most natural way to proceed with respect to the goals of counterfactual thinking, such as the construction of explanations.

Notice that when we use counterfactuals in explanation, we start from the consequent (the explanandum) and ask how the consequent C could have been different (C* been the case). Then we attempt to find an alternative A* to antecedent A such that it is true that "had A*

been the case, C* would have been the case". Tetlock and Belkin (1996, 20-21) find credibility in the suggestion that we give priority to this type of counterfactual reasoning. They point out that clear counterfactuals are such that is true that "had A* been the case, C* would have been the case, given everything else is held fixed to how it actually was". While this suggestion cannot be applied as restriction to all counterfactual reasoning (we can follow complex patterns of counterfactual dependencies in historical explanation, see Virmajoki 2022), it has an interesting feature from the interventionist point of view: If it is true that "had A* been the case, C* would have been the case, given everything else is held fixed to how it actually was" then A satisfies the definition of *actual cause* in interventionist framework (Woodward 2003, 77). Again, interventionism clarifies and explains an intuition behind discussions about historical counterfactuals.

It is also important to notice that an intervention might not be as plausible or likely as some "natural" change. For example, if there is an abandoned house with a barometer, it is more plausible that the barometer reading drops due to a storm than it is that someone changes the reading by using a finger; still only the latter would count as intervention. Likeliness or plausibility is irrelevant for what is the correct reading of the antecedent of the counterfactual. When considering a counterfactual scenario, we simply stipulate that antecedent took place due to an intervention. As Bunzl points out, "not all historically interesting counterfactuals need to be plausible. If our primary interest is in understanding what difference an event made, whether or not its occurrence is plausible doesn't really matter" (2004, 857).

There are strengths in this type of "antimetaphysical" strategy of using stipulations that go beyond our needs to explain. Sometimes it seems that the discussion on counterfactual histories is built on the assumption that there are possible events that were "really" the ones that would have taken place if remove something from the history – as if they were in line

waiting to take the place of the actual event once it was sucked out from the history with some metaphysical vacuum cleaner. Consider the following case. Robert Fogel's *Railroads and American Economic Growth* (1964) is widely considered a classic in counterfactual histories. As Bulhof summarizes,

"Using data from the time, and economic theories, Fogel constructed a mathematical model of the economic condition of the United States without the development of the railroad. This uses quantified rules of transformation derived from contemporary economic theories, and from them, Fogel argued for a modal conclusion: had the U.S. not developed the railroad, little would have been different" (1999, 166).

One of the leading critics of counterfactual histories, Richard Evans, writes as follows:

"Robert Fogel, in his analysis of the impact of railway building on the American economy, is not trying to imagine what America would have been like without railways, rather, he is building a statistical model of the US economy with the contribution of railways taken out, in order to show what the actual contribution of railways was. He is not, in other words, saying that the railways might not have been built. He is not constructing an alternative timeline. This is a fundamentally different procedure from the hypotheses of the true counterfactual historian, whose business is precisely to posit what might have been had some event or process not taken place." (2016, 462.)

It seems that Evans claims that counterfactuals related to causal explanation ("the actual contribution") are not what we should mean by counterfactual histories and that the two issues are separate. Counterfactual histories, it seems, must assume that the antecedent could have been the case in some very strong sense of "could". The final sentence is revealing: it seems that counterfactual histories must focus on removing something from the history and seeing what takes its place.

Notice that the last sentence goes against the normal way of formulating counterfactuals. We ask what would have happened, had some C been the case. We are not asking "What would have happened, had the actual C* not been the case". The difference is subtle. In the first case, we provide an exact starting point for a scenario; in the latter, we leave open a number of possibilities. It has been unlucky that counterfactual theories of causation were sometimes formulated in terms of the conditional "Had the cause not occurred, the effect would not have occurred" (see Lewis 1973). This leads to the question of what is meant by the cause not occurring. Currently, this type of ambiguity does not characterize counterfactual theories of causation (see Woodward 2003) but it seems that the ambiguity lives a life of its own in counterfactual histories.

Asking the question "What if the actual C* did not occur?" is a very unfortunate way of asking counterfactual questions. It suggests that we use some sort of a metaphysical vacuum cleaner to suck C* out of the history and see what event from the metaphysical space of possibilities fulfills the vacuum. This is not what is done when someone constructs a historical counterfactual. If it was, then we all should follow the critics of counterfactual histories. Yet, we do not need to follow the critics. We should stick to stipulating what kind of event we presume to have occurred due to an intervention and try to track the scenario through.

It is interesting to note that the kinds of problems that have been raised in historiography of human past does not seem to have a counterpart in the historical sciences. When we are asking what would have happened to the Cambrian explosion, had the Snowball Earth been different, we do not have similar tendencies to ponder about the most plausible way to change the Snowball Earth (or how they relate to human decisions, as there are none). Whereas, in the case of human history, one can ask even about the supposed changes in decisions whether they lead to a situation where the actors "cease to recognise or to acknowledge themselves as

the agents that they were" (Hawthorne 1991, 166; see Greene 2021) – and, therefore, lead to some unacceptable scenario – , one would never ask similar questions in the historical sciences. Surely, people may disagree in how the extent to which the Earth was a snowball would have affected how refugia would have been generated, but one would never ask whether some alteration in the antecedent, in this case, would have omitted the Earth its metaphysical essence.[6] It seems that our reasoning about historical counterfactuals in the case of the human past is confused by other dimensions of historical understanding. If counterfactuals were as complicated in their nature as the discussion about their use in human history suggests, then we should expect to find similar problems in neighboring fields. We do not find such problems, and therefore it would be best to stick to a straightforward understanding of the counterfactuals, such as the interventionist, and discuss metaphysical issues (such as when would a historical actor still recognizes herself) later and in other contexts.

5. Tracking Counterfactual Scenarios

This leads us to the final issue. How can we track counterfactual scenarios? We can set aside one common worry, that the way in which the antecedent occurs can make the scenario unintuitive. Suppose it is true that "Had Eddington made the decision not pursue the observation, the observation would not have been made". Now, one could object that, had Eddington made the decision, he would not have recognized himself anymore and lost his mind. This would have led other scientists to pursue the observation, as they would have been afraid that they have the same faith if they decline from pursuing the observation. The observation would have been made anyway, the argument concludes. We can set aside this type of surreal counterfactual by simply stipulating that the way Eddington decision is

[6] I thank an anonymous referee for pointing out the example.

supposed to be changed is independent of the rest of his psychological profile. We assume that Eddington continued to live his life as usual, the only difference being that he no longer pursued the observation. In other words, we assume that the change was due to an intervention, and we no longer need to worry about any consequences that the transition period itself might have on the outcome.

The question remains, how can we track counterfactual scenarios after the antecedent is specified? This is the second component if the minimal rewrite rule as formulated by Tetlock and Belkin (1996, 23-25). To be sure, there are different methods that an individual study can use. For example, we already saw how Fogel used statistical methods, data, and economic theories to answer what would have happened, had the railroads not been built. On the other hand, De Mey and Weber construct thought experiments that are arguments whose "premises … can be empirically supported, for example, by interviewing people or giving them questionnaires". They continue: "In general, thought experiments constitute a 'bottom-up' approach: we start from assumptions about how rational decision-making and/or nonrational psychological mechanisms determine the behavior of individuals in a social group. These assumptions can be tested by empirical but non-experimental techniques (interviews, questionnaires, text analysis). Then the assumptions are used to infer a causal relation on a higher level, the level of the population." (2003, 33-34.) Moreover, Tucker has argued that "counterfactual historiographic hypotheses are tested just like factual historiographic hypotheses … We suspend belief in a body of evidence that proves that the counterfactual never happened while maintaining all the rest of the evidence. The probability of the counterfactual depends then on the likelihood of the rest of the evidence given a counterfactual hypothesis." (2004, 229-230).

It seems that counterfactual scenarios require generalizations that are used to connect the antecedent to the later events. These generalizations can have many origins and forms, and it

would be important to give them (at least) the same attention in the philosophical analysis as has been given to the debate on the nature of transition periods. The problems are (i) what types of generalizations are needed, and (ii) what the relationship between the use of generalizations in the study of actual history and the study of counterfactuals is.

That historiography and historical sciences require generalizations is not a novel suggestion. It has been argued that generalizations are needed but these generalizations are not laws with a large scope. Rather, they are recurrent patterns of causal structure, with a much more limited scope of generalizability. The idea that limited generalizations provide explanatory understanding is widespread in the philosophy of science currently (see e.g. Jeffares 2008 on "localised regularities"). For example, as we saw above, Glennan's ephemeral mechanisms are short-lived and non-stable collections of interacting parts whose interactions can be characterized by *direct, invariant, change-relating generalizations* (Glennan 2010, 260). Also, Currie (2014, 2017) and Currie and Sterelny (2017) suggest that generalizations play an important role in historiographical explanations. Currie and Sterelny argue that narratives are constrained by our knowledge of causal structures of the past and that modeling practices can support narratives (2017, 5–8). Moreover, Currie (2014, 1169) is "inclined to see [even complex] explanations as leaning on a patchwork of regularities" and Currie (2017, 939) suggests that "historical scientists do not rely on mere empirical correlations to bring out dependencies; rather, they posit causal models that connect past entities."

Virmajoki (2023a) has suggested that counterfactual scenarios can be tracked once we make assumptions about the causal situation that held in the past and postulate principles that we assume guided the situation. Here we can note that the difficulty of the task varies greatly in different contexts:

In some (easy) cases, it is rather easy to track counterfactual scenarios. For example, the fact that an article inspired (and caused) a further article can be assessed by reading the latter article. We rely on the principle "An article comments on an earlier one only if its authors were influenced by the earlier article". Also, when Tetlock and Belkin (1996) illustrate the second component of the minimal rewrite rule (that the connecting principles and steps that we follow to consequent must be in accordance with the actual history) by saying that it is false that "had the Soviet Union not deployed missiles in Cuba, the United States would have invaded the island" because no such intention to invade was present – in fact, the idea of invasion was rejected due to the costs – they are relying on a principle that leaders are consistent in their actions and specific details of the situation. (Surely, this principle might be wrong but that is beyond the point here.) The second component of the minimal rewrite rule, therefore, is rather uncontroversial statement about our need to follow principles we find plausible and construct the steps in the contrafactual narrative in accordance to our principles.

In some (intermediate) cases, we need to do much more to establish a counterfactual dependency. For example, to establish the dependency of acceptance of a scientific idea on the age structure of a scientific community, we need to rely on (i) considerations of the credibility of Planck's principle that "[a new] scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it" (cited from Hull et al. 1978, 718) and (ii) on the prospects of applying that principle to the case at hand which requires demographic data and careful study of the views and discussions within the community.

In some (difficult) cases we need facts, non-trivial principles, and philosophical assumptions about the issues at hand. For example, when we study the dependence of the 20th-century developments in physics on the works of Einstein (theory) and Eddington (observation), we need to establish (i) facts (e.g. how widespread was the idea of spacetime-curvature in the

early 20th century), (ii) principles (e.g. how did international relationships in science work before and after the WW1), and (iii) philosophical assumptions (what is the role of observation in theory-acceptance?).

Moreover, we may notice that the problem of finding and justifying the relevant generalizations is made more severe by the fact that the antecedents commonly have many effects which have effects of their own. Tracking such chain of changes is extremely difficult. As Virmajoki (2023b, 91-92) notes, in many cases, we would require an enormous number of generalizations and, even if we had all the generalizations, we should still be able to see all sorts of contingencies that could affect the chain of events from the "outside" (or at least make strong and epistemically obscure assumptions about the insignificance of the contingencies). As Zhao (2023) argues, this is one of the main problem in counterfactual histories: while we can track individual statements, but narratives are a different beast.

This means that tracking counterfactuals is not easy and it rarely can lead to an agreement, as the principles and assumptions are quite difficult to assess (Virmajoki 2023a). The more general the principles used are, the more room there is for disagreement. However, counterfactual reasoning is possible and can be performed in a controlled manner. The commonly used term "thought experiment" does not do justice to the process. It is not about imagination but about applying generalizations that can be explicated and discussed. The evaluation of the generalizations also provides us with tools to identify wishful or otherwise biased counterfactual histories, the existence of which has been used as an argument against historical counterfactuals from time to time at least since Carr (1961, 97-98).

Finally, we can notice that all historiography requires generalizations and principles. As Lebow (2000) points out, even the actual history is reconstructed from evidence using generalizations and principles – also the reasonability of these works can be assessed on the

basis of the inferences. Surely, it is easier to establish a chain of events A-B-C-… in actual history where we can have direct evidence of each step. However, when we claim that the chain is a causal chain, we already suppose counterfactual dependencies and postulate some principles according to which A is the cause of B and so. The only difference between historiography of the actual world and counterfactual reasoning is in the fact that the inference chains are longer in the latter case. When we reconstruct the actual history, we make inferences from sources to past events and inferences about the causal relations between the actual events. In the case of counterfactual events, we make further inferences about what could have happened instead of the actual events. However, all the reasoning requires generalizations and principles. There is no guarantee that counterfactual reasoning is always more difficult than reasoning about actual history. For example, the claim that without Eddington, the observation of gravitational deflection would not have been made in 1919 seems more warranted than claims about what happened to the Roanoke Colony that disappeared at the end of the 1600th century. The former is grounded in well-documented historical events and the established significance of Eddington's contributions to astrophysics, making it a more concrete and evidence-based conjecture. In contrast, the mystery surrounding the Roanoke Colony, due to the lack of conclusive historical records and evidence, makes any claims about its fate more speculative and less verifiable.

6. Conclusion

In this article, we have seen what historical counterfactuals are, what their function is, and how we can reason about counterfactual scenarios. I have argued that historical counterfactuals have important functions that force us to resist skepticism towards their use. I also argued that there are confusions about how historiographical counterfactuals should be handled. The philosophical discussion has focused on how the antecedent of a counterfactual can be established. This has led to natural but ultimately unmotivated ideas about the

restrictions in acceptable antecedents in historical counterfactuals. I argue that we can simply stipulate the antecedent in terms of interventionism. Finally, I pointed out that counterfactual reasoning can proceed by using causal knowledge in the forms of generalizations and principles. While counterfactual reasoning is difficult and there are several places for disagreement in the process, the reasoning does not differ fundamentally from the more familiar types of historiography.

References

Beatty, John & Carrera, Isabel (2011). When What Had to Happen Was Not Bound to Happen: History, Chance, Narrative, Evolution. Journal of the Philosophy of History 5 (3):471-495.

Beatty, John (2016). What are narratives good for? Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 58:33-40.

Ben-Menahem, Yemima (1997). "Historical Contingency". Ratio 10 (2). 99–107.

Ben-Menahem, Yemima (2016). "If Counterfactuals Were Excluded from Historical Reasoning". Journal of the Philosophy of History 10 (3). 370-381.

Bulhof, Johannes (1999). "What if? Modality and history," History and Theory 38:2, 145–168.

Bunzl, M. (2004). "Counterfactual history: A user's guide". The American Historical Review 109. 845–858.

Carr, E. H. (1961). *What Is History?*. New York: Vintage.

Currie, Adrian (2014). Narratives, mechanisms and progress in historical science. *Synthese* 191 (6):1-21.

Currie, A. (2017). Hot-blooded gluttons: Dependency, coherence, and method in the historical sciences. *The British Journal for the Philosophy of Science, 68*, 929–952.

Currie, Adrian (2018). Rock, Bone, and Ruin: An Optimist's Guide to the Historical Sciences. The MIT Press.

Currie, A., & Sterelny, K. (2017). In defence of story-telling. *Studies in History and Philosophy of Science Part A*, 62, 14–21.

Daston, Lorraine (2009). ”Science Studies and the History of Science”. *Critical Inquiry* 35 (4). 798–813.

De Mey, Tim & Weber, Erik (2003). “Explanation and Thought Experiments in History,” History and Theory 42:1. 28–38.

Evans, Richards (2016). “Response”. Journal of the Philosophy of History 10 (3). 457-467.

Ferguson, Niall (1999) Virtual History: Alternatives and Counterfactuals. Basic Books.

Elster, John. 1978 Logic and Society: Contradictions and Possible Worlds. John Wiley & Sons.

Evans. Richards (2013). Altered Pasts: Counterfactuals in History. Waltham, Mass: Brandeis University Press.

Gallie, W. B. (1964). *Philosophy and the Historical Understanding*. New York: Schocken Books.

Glennan, Stuart (2010). Ephemeral Mechanisms and Historical Explanation. *Erkenntnis* 72 (2):251-266.

Gangl, G. (2021). Narrative explanations: The case for causality. *Journal of the Philosophy of History*. 15 (2):157–181.

Greene, Catherine (2021). Historical Counterfactuals, Transition Periods, and the Constraints on Imagination. *Hopos: The Journal of the International Society for the History of Philosophy of Science* 11 (1). 305-323.

Jeffares, B. (2008). Testing times: regularities in the historical sciences. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 39(4), 469-475.

Kohler, Robert, & Olesko, Kathlyn (2012). "Introduction: Clio Meets Science". *Osiris* 27 (1). 1-16

Lebow, R. N. (2000). "What's so different about a counterfactual?" World Politics 52. 550–585.

Lewis, David (1979). Counterfactual Dependence and Time's Arrow. *Noûs* 13 (4):455-476.

Lewis, David K. (1973). *Counterfactuals*. Cambridge, Blackwell.

Maar, Alexander (2014). "Possible uses of counterfactual thought experiments in history," Principia: An International Journal of Epistemology 18:1. 87–113.

Maar, Alexander (2016). "Applying D. K. Lewis's Counterfactual Theory of Causation to the Philosophy of Historiography". Journal of the Philosophy of History 10 (3). 349-369.

Nolan, Daniel (2016). "The Possibilities of History". Journal of the Philosophy of History 10 (3). 441-456.

Hull, David L. & Tessner, Peter D. & Diamond, Arthur M. (1978). Planck's Principle. *Science* 202 (4369). 717-723.

Reiss, Julian (2009). "Counterfactuals, Thought Experiments, and Singular Causal Analysis in History" Philosophy of Science 76 (5):712-723.

Rosenfeld, Gavriel (2016). “The Ways We Wonder “What If?” Towards a Typology of Historical Counterfactuals”. Journal of the Philosophy of History 10 (3). 382-411.

Soler, Léna & Trizio, Emiliano & Pickering, Andrew (2015). Science As It Could Have Been. Discussing the Contingency/Inevitability Problem. University of Pittsburgh Press.

Stanley, Matthew (2003). ‘An Expedition to Heal the Wounds of War’ The 1919 Eclipse and Eddington as Quaker Adventurer. *Isis* 94 (1). 57–89.

Stanley, Matthew (2019). *Einstein's War: How Relativity Triumphed Amid the Vicious Nationalism of World War I*. Dutton

Sunstein, Cass (2016). “Historical Explanations Always Involve Counterfactual History”. *Journal of the Philosophy of History* 10 (3). 433-440.

Tetlock, Philip E, and Aaron Belkin (1996) “Counterfactual Thought Experiments in World Politics: Logical, Methodological and psychological Perspectives.” In Tetlock, Philip E, and Aaron Belkin (eds.) *Counterfactual Thought Experiments in World Politics*. Princeton University Press

Tucker, Aviezer (2016). “Historiographic Counterfactuals and the Philosophy of Historiography. An Introduction”. Journal of the Philosophy of History 10 (3). 333-348.

Tucker, Aviezer (2004). *Our Knowledge of the Past: A Philosophy of Historiography*. Cambridge University Press.

Virmajoki, Veli (2018). Could Science Be Interestingly Different? *Journal of the Philosophy of History* 12 (2).

Virmajoki, Veli (2022). What Should We Require from an Account of Explanation in Historiography? *Journal of the Philosophy of History* 16 (1). 22-53.

Virmajoki, Veli (2023a). “Frameworks in Historiography. Explanation, Scenarios, and Futures”. Journal of the Philosophy of History.

Virmajoki, Veli (2023b). *Causal Explanation in Historiography*. Springer Nature Switzerland.

Weber, Max (1949). On the Methodology of Social Sciences. The Free Press of Glencoe.

Woodward, James (2003). Making Things Happen. A Theory of Causal Explanations. Oxford University Press.

Woolf, Daniel (2016). “Concerning Altered Pasts: Reflections of an Early Modern Historian”. Journal of the Philosophy of History 10 (3). 413-432.

Zhao, Helen (2023). Counterfactual History: Three Worries and Replies. *Journal of the Philosophy of History* 17 (1):9-30.