Counterfactuals and Causal Explanation in Historiography

Päivi Seppälä (013226554)

Pro gradu, 4.9.2012

University of Helsinki

Department of Philosophy, History, Culture and Art Studies

Contents

1 Introduction

If there is such a thing as a sense of reality – and no one will doubt that it has its raison d'etre – then there must also be something that one can call the sense of possibility.

– Robert Musil

In his novel *the Man without Qualities*, Robert Musil contemplates insightfully the human ability to consider possibilities and to imagine how things might have happened or would happen. Since the ability to think in modal terms is a universal feature of human cognition, it should not come as a surprise that not even academic research is immune to this way of thinking. *Counterfactual questions* on past possibilities tend to tingle the imagination of a professional historian in the same way as they tingle the imagination of any of us – it can be hard to resist asking: what if this and that had been different? It has been doubted, however, whether the human ability to think about the past possibilities should have any methodological or epistemic role in the work of a rigorous history scholar. Should a self-respecting historian stick to her sense of the actual past and save the disturbing *what if*-questions for the occasions of playing "parlour-games with *might-have-beens*", as the distinguished methodologist E.H Carr (1964: 97) suggests?

This thesis deals with the use of counterfactuals in historiography[1] from the point of view of causal explanation. It studies the various roles that counterfactuals and counterfactual reasoning play in causal explanation and causal thinking in historiography. The aim is to defend the use of counterfactuals in the context of causal explanation and to create methodological guidelines for that use.

The motivation for the topic stems from the heated debate that has surrounded the use of counterfactuals in historiography. In this debate, the use of counterfactuals has been defended on the basis of claiming that causal thinking and causal explanation are inherently linked to counterfactual reasoning. This claim has not, however, been fully argued for. There exists a need in the philosophy and methodology of historiography to explicate in what ways causal explanation and counterfactual

[1] I will use the term 'historiography' to refer to the research on the human past. In order to avoid confusions the term 'history' is reserved to denote that past, not the activity of studying it (see e.g. Tucker 2011).

reasoning are linked together. This is the point where philosophy of science and its theories of causal explanation step in.

The theory that is used in this thesis to explicate the role of counterfactuals in causal explanation is called *the contrastive counterfactual theory of causal explanation*. Both Petri Ylikoski (2001; 2007) and James Woodward (2003) have introduced accounts that can be regarded as versions of this theory. In the context of Woodward's account it is also common to talk of *interventionist* or *manipulationist theory* of causation and causal explanation. The motivation to apply the contrastive counterfactual theory to the debate on counterfactuals comes from two sources. First of all, the theory has become the standard theory of causal explanation in philosophy of science within the past decade. In spite of this, there has been relatively little discussion of this theory in the context of historiography. It is, therefore, high time to correct the situation with an extensive discussion about the implications of the contrastive counterfactual theory to historiography. Is the theory able to provide some methodological fruits that historians can pick when seeking for the causes of historical events? James Woodward (2003: 313–314) has doubted the applicability of his theory to other areas of historiography than economic history in few of his remarks. Woodward's doubts give further motivation to choose the contrastive counterfactual theory as the theory to be discussed and applied. Is Woodward right in his doubts? Is causal explanation in historiography beyond the reach of the contrastive counterfactual theory?

The thesis consists of five discussion chapters. The second chapter introduces the basic concepts of the thesis such as 'counterfactuals' and 'causal explanation' and analyzes the debate of counterfactuals that has churned among historians within the past decades. In the third chapter, the basic ideas of the contrastive counterfactual theory are jotted down. Chapters four, five and six are dedicated to applying the contrastive counterfactual theory to the most puzzling questions in philosophy of historiography in order to see how far the theory can go in solving these issues. Most of the topics discussed are illustrated with examples taken mainly from political history, social history, history of international relations, and economic history. However, the arguments and conclusions should be applicable to all fields of historiography. The final closing chapter seeks to give an answer to those critics who question the use of counterfactuals in historiography, and to provide a toolkit for counterfactual explanatory reasoning in historiography based on the findings of the previous chapters.

2 The debate on what-might-have-beens

The question in history is never what must, or what might have taken place, but solely what the evidence obliges us to conclude did take place.

– Michael Oakeshott (1966[1933]: 139)

The purpose of this chapter is to lay the background to the following chapters by presenting and analyzing the methodological discussions on historiographical counterfactuals that philosophy of historiography has witnessed during the last century or so. These discussions show that not all historians are willing to regard counterfactual questions as serious research questions. Some – I will call them *antagonists of counterfactuals* – even want to weed out counterfactual reasoning from historiographical research as a methodologically flawed enterprise (Oakeshott 1966[1933]: 139; Carr 1964: 96–97; Croce 1966: 557; Fischer 1971: 18). On the other hand, *the protagonists of counterfactuals* are quick to point out that even the toughest opponents of counterfactuals implicitly engage in counterfactual reasoning in their historiographical studies (Bulhof 1999; Kaye 2010; Tetlock & Parker, 2006: 18). The dividing line between the antagonists and protagonists seems to be extremely polarized.

Next I will examine, analyze and evaluate the arguments presented by both sides of the controversy and try to locate where the polarization stems from. The analysis of the antagonists' arguments brings up the worries and problems that any account defending the use of counterfactuals needs to answer. This chapter provides some preliminary answers, but the reader is asked to be patient, and wait for the summaries of the concluding chapter. The final answers to the antagonists' arguments can be given only after careful scrutiny of a theory that justifies the use of counterfactuals as a methodologically sound enterprise. This theory is presented in chapters 3, 4, 5 and 6. The analysis of the protagonists' arguments reveals that the protagonists also need a theory – namely, an explicit theory of causal explanation to strengthen their arguments against the antagonists. Before discussing these issues, however, let's have a look at the most central concepts of the thesis.

2.1 Counterfactuals and counterfactual reasoning

Counterfactual reasoning is no new game in the town of historiography. Tacitus mourned the early death of Germanicus as follows: "Had he been the sole arbiter of events, had held the powers and title of King, he would have outstripped Alexander in military fame as far as he surpassed him in gentleness, in self-command and in other nobel qualities" (quoted in Tetlock & Belkin 1996: 3). There is no lack of examples if one wants to elaborate what kind of claims historians make when immersing themselves in counterfactual reasoning. Here are a few more examples:

- If Themistocles had not planned and led the Athenians' attack, the Greeks would have lost the battle of the Salamis in 480 BC (Hanson 2006).
- Had there been no railroad connections, it would not have significantly restricted development of the American economy in the last half of the nineteenth century (Fogel 1964).
- If Kennedy had shown more resolve prior to the Cuban missile crisis, Khrushchev would not have deployed the missiles (Lebow & Gross Stein 1996).

These examples reveal the general idea of counterfactual claims: *a counterfactual claim is a subjunctive conditional in which the antecedent is known or supposed for the purposes of argument to be false, but in which the consequent is supposed to follow (not logically!) from the antecedent* (Chisholm 1946: 289–290).

Counterfactual reasoning is the term I use to refer both to the activity of formulating and justifying counterfactual claims and to the way of formulating research questions in a counterfactual form. When one formulates and justifies counterfactual claims, one makes *counterfactual inferences*. Historians quite often approach counterfactuals by entertaining *counterfactual questions* starting with "what if" or "what would have happened". For instance, one can ask: What if Stalin had been ousted as general party secretary of the communist party? What would have happened to the development of the Soviet Union in the 30's? Pure definitions, however, do not tell what the point of presenting these questions is. Should one or should one not engage in counterfactual reasoning in the first place?

2.2 What makes counterfactuals dubious?

> When judgment is brought to bear upon a fact, the fact is taken as it is and not as it might otherwise have been… Historical necessity has to be affirmed and continually reaffirmed in order to exclude from history the "conditional" which has no rightful place there. What is forbidden is… the anti-historical and illogical "if". Such an "if" arbitrarily divides the course of history into necessary facts and accidental facts… This is a game which all of us in moments of distraction or idleness indulge in, when we muse on the way our life would have turned out if we had not met a certain person… When the attempt is made to play this sort of game on the field of history, where it is thoroughly out of place, the effect is too wearisome to be maintained. (Croce 1966: 557.)

The above comment by Benedetto Croce nicely reveals some of the general arguments that the antagonists of counterfactuals have presented. Croce's style of argumentation, with its richness and rhetorical power, also represents quite nicely the way of writing of many of the 20th century antagonists of counterfactuals whose main message has been the total rejection of counterfactual reasoning. Croce's main message is blunt and clear: historians should concentrate on historical facts that Croce labels as "necessary facts". Counterfactual thinking, on the other hand, is methodologically disastrous and should be restricted to the contemplation of the coincidences of one's personal life. In what follows, I will group the arguments of the antagonists of counterfactuals into four categories: *1) Metaphysical arguments, 2) Semantic arguments, 3) Epistemological arguments 4) Arguments based on disciplinary identity.*

2.2.1 Metaphysical arguments

Croce's outburst against counterfactuals reveals one kind of metaphysical commitment that a historian might have: the idea of historical events as necessary facts that could not have turned out any other way than they did and the methodological idea that this metaphysical commitment should reverberate to the actual research one conducts. In general, there have been two lines of argument in the methodological literature which can be characterized as metaphysical:

(M1) Counterfactual reasoning leads the historian to consider the obscurities of the ontology of possible worlds (Hawthorn 1991: 5–6; Elster 1978: ch. 6). Thus, in order to have a methodologically sound basis for the use of counterfactuals, we need a philosophically and conceptually valid ontology of possible worlds.

(M2) Counterfactual reasoning lures one into thinking about the age-old riddles of determinism, contingency, necessity, fate or free will (Fischer 1971: 18; Croce 1966: 557; Carr 1964: 96–97).

The long tradition of philosophy of history and historiography has taken it almost as a given that a writer with the aim of developing sophisticated methodology of historiography, should take a stand on the issues of metaphysical necessity, contingency, possible worlds or free will[1]. At least this is the conclusion one can infer from the abundance of methodological literature filled with these questions, although these questions are sometimes taken up only to dismiss them as making no difference to the actual work of an historian (e.g. Carr 1964 ch. 4 and 5; Hawthorn 1991: 5–18; Ferguson 2011[1997]).

It is true that counterfactual claims are linked to a vast bunch of metaphysical problems such as the ontology of possible worlds and the problems of modal concepts like 'necessity' and 'contingency'[2]. Thus, the question to be solved is whether these metaphysical issues rise up in the context of empirical historiographical research: are counterfactuals used in historiographical research in such a way that the working historian is compelled to take a stance in these metaphysical issues? As the arguments of the next few paragraphs will reveal, I am most willing to go along with E.H Carr (1964: 96) who suggests that a historian could do without the "inevitable" as a metaphysical doctrine: metaphysical questions should be left to poets and metaphysicians since these questions have no a direct effect on the actual work of a historian. What I do not agree on is Carr's (ibid.) idea that counterfactual historians have introduced tough and obscure metaphysical themes into historical research that should be about factual descriptions and causal explanation that has, according to Carr, nothing to do with counterfactuals. Therefore, it will be argued that counterfactuals are a way to analyze historical causal claims and historical explanations – and this analysis need not bear any immediate inherent connection to metaphysical questions of determinism and contingency, let alone to possible worlds. Let me elaborate on this.

[1] This is not surprising, since, historically speaking, *philosophy of historiography* has not been separate from *philosophy of history* that – unlike contemporary philosophy of historiography - encompasses the metaphysical issues of the nature of time and history and asks whether history is progressive, continuous or discontinuous etc. (Vašiček 2011).

[2] For further discussion on the ontological status of the possible worlds see, for instance, Lewis (1986a), and Melia 1998. Kripke (1972) is a classic study on the concept and semantics of necessity. For a general overview of epistemology of modalities, see Vaidya (2011).

First of all, I want to tackle the question of possible worlds. In philosophy of science, the idea that causal explanation and theory of causation have something to do with possible worlds stems from David Lewis's work (1973a; 1986b; 2000). However, the issue of possible worlds has not been frequently brought up in methodological discussions of historians[3]. The reason for this might be that the theme of possible worlds is relatively new in the philosophy of science. It might also simply be that historians shun Lewis's theory due to its extreme richness in philosophical and technical detail. In any case, the contemporary discussion of possible worlds has some of its roots in Lewis' theory of causation and causal explanation and his idea that the truth conditions of counterfactuals should be stated in terms of similarity conditions between possible worlds.

Lewis defines causation as an ancestral of counterfactual dependence in the following way: *c causes e* if only if there is a chain c, d_1 … d_n, e such that each member of the chain (except c) is counterfactually dependent on the event preceding it. Crudely put, *c* causes *e* if and only if i) if *c* occurs, *e* occurs ii) if *c* did not occur, *e* would not occur. These counterfactuals are deemed as true sentences, if i) the worlds where *c* occurs and *e* occurs are more similar, and therefore closer to the actual world than the ones where *c* occurs and *e* fails to occur, ii) the worlds where *c* and *e* do not occur are more similar to the actual world than the ones where *c* does not occur and *e* occurs.

Lewis' original theory of causation and causal explanation has been shown to face several severe problems – it fails to secure the temporal asymmetry of causal dependence, and it solves neither the problem of transitivity or the problem of pre-emption (Menzies 2009). When these problems are viewed in the light of the fact that Lewis' theory explicitly rests on the idea of possible worlds that carry their own baggage of ontological problems, one could indeed draw a desperate conclusion: dealing with counterfactuals leads to complicated conceptual problems, and in the end, to severe metaphysical and ontological problems concerning the ontological status of possible worlds. Luckily, as the next chapter will reveal, Lewis's analysis is not the only game in the town of counterfactual theories of causal explanation.

The idea of possible worlds does not need, however, lead to big ontological questions of the existence of ontological worlds, although Lewis himself is also known for his modal realism. According to Lewis, possible worlds actually exist. As Kripke

[3] See, however, Climo & Howells (1976); Elster 1978: ch. 6; Bulhof (1999: 167); Fearon (1991: 176); Hawthorn (1991: 21–25) Tetlock & Belkin (1996: 18), Nolan (2011: 10).

(1972: 44) points out, possible worlds can be thought of as stipulations and invitations to imagine some things being different from the way they are (or were) in the actual world. To Kripke, possible worlds are not actual entities or distant countries that we find looking through a (metaphysical?) telescope. This would seem to be a reasonable attitude to adopt in any empirical field of study where the causal claims are based on evidence and theoretical considerations *concerning this world*. This does not mean that one could not use the idea of comparing the similarities between possible worlds as a conceptual tool when one tries to formulate and assess one's causal claims. One could, for instance, say that based on the evidence, it seems that the world where Franz Ferdinand gets shot on June 28th in Sarajevo and where the First World War breaks out is closer to the actual world than the one where Franz Ferdinand gets shot and the World War one does not occur. But it is very important to notice that it is the empirical evidence that justifies the similarity comparisons between the worlds and in the end the truth of one's causal claims. *The possible worlds that one speaks about do not do provide any evidence of their own – they are tools used to assess what kind of claims the evidence can support.*

Even if it is quite easy to avoid the ontological problems related to possible worlds, one can still ask whether Lewisian possible world semantics is the most useful tool to approach historians' counterfactual reasoning. The volume *Counterfactual thought experiments in World Politics* points strongly towards the direction that it is not. The editors Philip E. Tetlock and Aaron Belkin (1996: 18) report in a footnote that they had asked the contributors of the volume to perform a Stalnaker-Lewis test on the historical cases they were studying. In other words, the contributors were asked to use the Lewisian semantics of possible worlds in order to assess the truth of counterfactual claims they were proposing based on their work, for instance, on the Soviet politics or U.S policy towards Iran. Tetlock and Belkin, however, report that none of the contributors was successful in implementing the Stalnaker-Lewis test in their cases. The empirical evidence and theories that the contributors had, did not simply give them enough causal knowledge to stipulate possible worlds and to do similarity comparisons between these worlds and the actual world[4].

[4] As comparison the other tests that the authors were asked to do included evaluating their counterfactuals in terms of their clarity, projectability and their consistency with historical facts, theories and statistical generalizations. The authors found these tests much more useful in assessing their counterfactuals.

Another big metaphysical question that counterfactuals seem to give rise to is the question of determinism: do things happen the way they do by necessity or is there real contingency in the world? E.H. Carr (1964: 96–97) is right in pointing out that historians who explicitly frame their research questions in counterfactual terms seek to "measure" the contingency of the historical event in question. But, is this metaphysically so dubious that it should be left to poets, as Carr ends up suggesting? Not necessarily. For instance, when Robert Fogel (1964) asks the question of the significance of railroads for the economic growth in the U.S. he is definitely not making a case for any metaphysical doctrine of the world history and historical progress. Instead, he is interested in knowing whether railroads were the only route to the economic prosperity of the U.S. or whether there were other historically possible routes that might have resulted in the similar level of GDP. In other words, Fogel is trying to assess how big a role the other factors played in fixing the outcome: if we take railroads out, would the outcome still be the same? Chapter 6 further deals with the question of how to approach the issues of contingency and necessity without waking up the metaphysical boogieman.

2.2.2 Semantic arguments

One way argue against counterfactuals is to claim that counterfactuals lack clear meaning or that their truth conditions are too vague. James Woodward (2003: 122) introduces this quite popular argument by bringing up the following slightly humorous example by Quine:

(SA) If Julius Caesar had been in charge of the U.N. Forces during the Korean War, then he would have used (a) nuclear weapons or (b) catapults.

It is indeed quite hard to pinpoint on what basis one could decide which one of the consequents, (a) or (b), is correct. It can also be quite hard to imagine Caesar in the context of the 20th century world politics. But one can still ask whether this example works as a good counterexample against counterfactuals that are actually used in historiographical research. Are the counterfactuals I presented in the introductory chapter of the thesis on the same level of ridiculousness than Quine's counterexample to counterfactual? My answer is negative.

In the last couple of decades, some historians propagating counterfactuals have tried to establish counterfactual history as a serious methodological movement. In order to avoid counterexamples like the one by Quine, they have been eager to stress that counterfactuals should be *well-defined* (Tetlock & Belkin 1996: 19–21). They insist that in *well-defined counterfactuals* the antecedents should be historically plausible in the light of the evidence, and that there should always be a connection between the antecedent and consequent that one is justified to infer from the evidence and theoretical principles. According to this view, Quine's example is implausible because we cannot use historiographical evidence to place Ceasar in the context of the 20th century world politics and to tell how he would have reacted. As it will be shown later, there can be well-defined counterfactuals in historical research, and therefore, it would be quite erroneous to regard Quine's example as a universal dismissal of counterfactuals. Quine's example only shows that ridiculous and murky counterfactuals are what they are: ridiculous and murky. We should, thus, come up with a theory that helps to distinguish good counterfactuals from bad ones. The following chapters of the thesis, chapter 5 in particular, will also take up the question to what extent it is possible to clearly separate good counterfactuals from bad ones.

2.2.3 Epistemic arguments

The epistemological argument stems from the following idea: "The field of the historian is, and must long remain, the discovery and recording of what actually happened" (Burton Adams 1909: 236). Burton Adams's comment is based on the presupposition that historians are epistemically justified to use the evidence available to them to make inferences that result in interpretations of what happened. This is a presupposition that most working historians and philosophers of historiography agree on. The question to be raised, however, is whether it is epistemologically and methodologically justified to to use evidence to infer things that did not happen. Michael Oakeshott (1966[1933]: 139) does not seem to think so when he states: "The question in history is never what must, or what might have taken place, but solely what the evidence obliges us to conclude did take place".

Oakeshott draws a strict epistemological distinction between the "factual" inferences and "counterfactual" inferences. Factual inferences are those inferences that concern the things that happened and the things that can be known. Factual inferences are epistemically respectable since there is a causal link between the past events and the

evidence we now have: the current evidence consists of the traces that have resulted from the past events and that still remain. The causal link between the past and the current evidence justifies the use of that evidence for the purposes of inferring what happened in the past.

According to Oakeshott, counterfactuals, on the other hand, concern alternative worlds or alternative happenings that did not take place in the past. These are things that the evidence necessarily stays silent on – there are no direct causal connections between our evidence and these alternative worlds or the things that did not happen. I believe that this epistemological distinction between factual and counterfactual motivates those writers who dismiss counterfactual reasoning as mere flight of fancy, wishful thinking or at best pure speculation. Counterfactual inferences can simply never have the same epistemological status than factual claims, since evidence does not support counterfactual inferences in the same way it supports factual claims.

The principled epistemological objection to counterfactuals can be shortly formulated as follows:

(PEO) The evidence only "tells" us about the things that did happen. Since counterfactuals do not concern things that happened – the antecedents and consequents are by definition things that did not happen – they do provide any knowledge of past events, and thus do not belong to historiographical research.

As already stated, the objection is based on the assumption that our evidence can only tell us about the things that the evidence is causally linked to, that is, the things that have actually happened in the past. This is the basic assumption of all empirical fields of study. There is simply no causal link to things that did not happen, and therefore our empirical fields of study do not concern the things that did not happen. Thus, the question is whether the antecedents and consequents of counterfactuals refer to things that have actually taken place or do they refer to alternative worlds where we have no causal and, therefore, no epistemic access. If counterfactuals refer to things that we have no epistemic access to, I am most willing to agree with the antagonists of counterfactuals, and claim that counterfactuals on past events are nonsense. If, on the other hand, our counterfactual claims somehow are linked to things that our evidence can "speak of", the protagonists of the counterfactual reasoning have found the way to escape the principled epistemological objection.

There is, indeed, a way to escape the principled epistemological objection: it is to argue that counterfactuals are conceptually linked to causal claims that speak of the actual causal relations of the past. So, if we have evidence on these causal relations, we have at least some kind of evidence of the related counterfactuals. This line of argumentation will be discussed together with the principled epistemic objection in chapter 4.

Even if there was a way to avoid the principled epistemic objection, it would not clear counterfactuals from all the epistemic doubts, however. Rejecting the principled epistemic objection means that our evidence can be used to determine, *in principle*, the truth of counterfactual claims. The question still remains whether this is possible *in practice*. It might be that it is more difficult to use the evidence to support counterfactual inferences than to answer, for instance, *what happened*-questions. This objection is also a valid point, since we have good grounds to think that answering simple *what happened*-questions is not always straightforward, either.

First of all, answering *what happened*-questions requires a lot inferential work, since historians' evidence, the traces from the past, does not directly "tell" us what happened: *one can infer what happened only on the basis of reliable evidence.* This means that one needs to make sure that a certain piece of evidence can be reliably used as evidence for the purpose of making inferences on a past event. Arriving at conclusions of the reliability of evidence is not straightforward inferential procedure: it requires getting answers to causal *why*-questions. In order to assess the reliability of the evidence, one usually needs to ask: Why is this evidence the way it is? What is the causal story behind it? Are there pieces of evidence that seem to tell a conflicting story of the same event, and if so, why? For instance, internal source criticism requires that when dealing with documentary evidence one asks why this document was created in the first place. Why did the author of the document write in this way instead of writing in some other way? Only by answering these questions one can infer what kind of things a piece of document can be evidence for, whether the evidence should be used to shed light on the thoughts and values of its author, whether it reveals something interesting of the social context of its time, or whether the document could be also treated as a reliable "eye-witness report" of a particular event.

Answering *what happened*-questions is not complicated only by the complexity of inferences related to assessing the reliability of evidence, but also *the chronic lack of evidence* that historians constantly face. In other words, historians face the problem of

underdetermination not only in principle, but also very often in practice. There are cases where we do not have enough traces from the past that could be used to separate between competing descriptions of what happened (*see* Turner 2007 for further discussion). This might be due to the fact that the relevant evidence has been destroyed at some point in history or due to the fact that the past event of interest left little or no traces to begin with. The fact that there is not always enough evidence even to decide what happened justifies asking whether we can ever have enough evidence to support counterfactual claims. Based on these considerations one can formulate *the modest epistemic objection* as follows:

(MEO) Since the chronic lack of evidence makes reconstructing the past events hard enough in itself, how could we even dream of having strong evidential support for counterfactual inferences? Due to the chronic lack of evidence we should not regard constructing and justifying counterfactuals as the most essential part of the work of a historian.

Since the antagonists have mainly concentrated on dismissing counterfactual reasoning as mere flight of imagination and "a parlour-game with might-have-beens" (Carr 1964: 97), I have not come across this kind of modest epistemological argument in the comments of antagonists. David Hackett Fischer (1971: 19) is an antagonist, who comes rather close to presenting the modest epistemological argument when he criticizes Robert Fogel's way of using counterfactuals. Nevertheless, Fischer still rejects the use of counterfactuals for principled reasons: "It is always possible to convert any historical problem into nonhistorical one, but why should a scholar go out of his way to make a difficult problem impossible? History is tough enough as it – as it *actually* is" (ibid.). While the antagonists have been quite silent on the actual problems that historians face when using counterfactuals, the protagonists of counterfactuals who explicitly engage in counterfactual reasoning in their studies have been willing to admit the problems that the lack of evidence poses for the use of counterfactuals (e.g. see Lebow & Gross Stein 1996). The modest epistemological objection will be addressed again in chapter 5 that discusses the use of evidence in supporting counterfactual claims.

2.2.4 Arguments based on disciplinary identity

"What is forbidden is… the anti-historical and illogical 'if' "(Croce 1966: 557). Croce's statement is a representative example of the demarcationist tendency of many antagonists of counterfactuals. To declare something "anti-historical" or "a-historical" is a powerful rhetorical tool that is usually used as a knockdown argument to nail the coffin of counterfactual reasoning for good. Where does this kind of rhetoric originate from?

Counterfactuals have aroused strong emotions for and against throughout the whole 20th century (Kaye 2010). This might explain why methodologist have been keen on making hasty categorical distinctions between "factual" and "counterfactual" reasoning and also between "history" and "anti-history". The eagerness to jump into drastic either-or conclusions has led to the lack of balanced and conciliatory argumentation between the opposing views. Luckily, after the beginning of the 1990's and the publication of Geoffrey Hawthorn's monograph *Plausible Worlds*, a slightly more modest methodological discussion on counterfactuals has emerged.

In the arguments of antagonists, history is contrasted with "anti-historical" counterfactual reasoning in two ways. Usually the contrast is drawn between the rigorous history writing that concentrates on the things that actually happened and the flights of fancy or emotional irrationality that the counterfactual historians are accused of. In this case, the demarcation between history and "anti-history" is done between an epistemically respectable systematic field of study and epistemically dubious flights of fancy that are not based on evidence, but rather on imagination and emotions.

Another demarcation line has been drawn between history as *a humanistic study* or *a study of man* and *the natural science* (and also social sciences if the social sciences are considered to imitate the methodologies of natural sciences). Oakeshott (1966 [1933]: 129) states: " The distinction … between essential and incidental events does not belong to the history at all: it is a monstrous incursion of science into the world of history". Counterfactual reasoning is thus seen as epistemically respectable, but only in the context of science. If history is seen as distinctively different from the sciences, not only in respect to the object of study but also methodologically, all the methodological tools that are believed to be imported from the sciences are treated with doubt or straightforward rejection. The debate of the 1970's on Robert Fogel's work on the construction of the U.S railroads and the U.S civil war and the abolition of slavery was

partly motivated by the strong demarcation between the traditional historiographical methodology and the (social) scientific methodologies that the *cliometric or scientific movement* in economic history was promoting. The movement, and Robert Fogel, in particular, regarded the use of counterfactuals as central to social scientific causal reasoning whereas the proponents of the "traditional" doubted the applicability of this type of reasoning in the context of historiography (Fogel & Elton 1983).

Arguments of disciplinary identity are rhetorically strong, but drastic either-or distinctions can quite often turn out to be *descriptively weak.* Many proponents of counterfactuals have claimed that this observation also applies to the case of counterfactuals in historiography. They have pointed out that almost all historians, even the strictest objectors, are relying on counterfactual reasoning in one way or another (Bulhof 1999; Kaye 2010). According to Tetlock & Parker (2006: 18), the use of verbs such as *influenced, shaped, responded, transformed, triggered, molded, precipitated* takes the historian into the realm of causal claims, and if protagonists are right straight to counterfactual reasoning. The claim that every historian is, in fact, in one way or another engaged in counterfactual reasoning is an apt way of pointing out that no one is actually practicing what the high-fly rhetoric of disciplinary identity is preaching. However, showing that there is a practice cannot be taken as a justification for why there should be such a practice. One needs to be able to explain why it is epistemically and methodologically justifiable and even desirable to engage in counterfactual reasoning. The following section will deal with this issue of justification.

2.3 Why are counterfactuals indispensable for a historian?

Throughout the 20th century, counterfactual reasoning has had its supporters (for a general historical overview, see Kaye 2010). The most common general methodological argument for counterfactuals has been that counterfactuals are indispensable for identifying causes, and hence, for explaining events in history (Weber 1949: 164–188; Aron 1961: 156–186; Fogel 1964; Hawthorn 1991, cp. 2; Fearon 1991, 1996; Tetlock & Belkin 1996; Stanford 1998: 85–93; Bulhof 1999; Bunzl 2004; Lebow 2007; Megill 2007: 100, 155; Kaye 2010). Despite the prominent role given to the arguments that link causal explanations to counterfactuals, there are several other arguments presented to justify the explicit use of counterfactuals in historiography. Thus, before going into a closer analysis of the arguments on causal explanation, I will briefly introduce Daniel Nolan's (2011) list of eight reasons for why historians should use counterfactuals. The

list highlights the numerous methodological gains that the explicit use of counterfactuals brings with it.

2.3.1 Eight reasons for using counterfactuals

Daniel Nolan (2011) has provided the most systematic overview of the arguments that various proponents of counterfactuals, notably Richard Ned Lebow (2007; 2010), Tetlock and Belkin (1996) and Tetlock & Parker (2006), have presented. Nolan does not claim that his list is an exhaustive list of all the reasons why counterfactuals are important in historiography, but in any case Nolan's suggestions can be said to provide a good starting point for a more refined research on the matter. Nolan lists in total eight uses of counterfactuals that he believes to be justified on the basis of the epistemic concerns of historians and the ways in which historiographical research is conducted. Four uses are of the sort that one does not need to know the truth of the counterfactuals when discussing them. Instead, the other four require, according to Nolan, the ability to assess the truth of the counterfactuals one is dealing with.

Uses of counterfactuals that do not require knowing the truth-values of counterfactuals

1) The use of counterfactuals stimulates historians' imagination and suggests new hypotheses to be tested. For instance, when doing a comparative case study on the feudalisms of the Japanese shogunate and Medieval Latin Europe, it might be useful to ask how the Japanese shogunate would have dealt with the split of religious authority. The answer to this question helps to highlight the differences and similarities of the feudalists systems. (Nolan 2011: 4)
2) Comparing the counterfactuals that the different sides of a historiographical controversy are willing to accept clarifies what the debating sides are in fact disagreeing on in their differing causal accounts: the other side might claim that the absence of this and that factor would have caused the result of interest while the other side strongly disagrees on this and claims that if the suggested factors had not been present, the result would still have occurred (Nolan 2011: 5).
3) The explicit use of counterfactuals helps to tackle the so called *hindsight bias*, the human tendency to see past events as more inevitable or overdetermined than they actually were considered to be when they occurred (Nolan 2011: 5; *see also* Lebow 2007; Tetlock and Belkin 1996: 15).

4) Describing the thoughts, beliefs and desires of historical actors might sometimes require that one is interested in the counterfactual reasoning that the actors themselves engaged in when making decisions (Nolan 2011: 6–7).

Uses of counterfactuals that require assessing their truth-values

5) Thinking in counterfactual terms and trying to assess the truth of counterfactuals is interesting in its own right (Nolan 2011: 8–9).
6) Counterfactuals are closely connected to causation, and causal reasoning sometimes seems to require counterfactual considerations (Nolan 2011: 9–14).
7) Counterfactuals are needed in causal explanations: we take those factors to be explanatory without which the event to be explained would not have occurred (Nolan 2011: 14–15).
8) Counterfactual information is needed in assessments of moral responsibility: we need counterfactual information in order to appraise or blame historical actors. The attributions of moral responsibility partly depend on considerations whether an outcome that an actor is blamed or appraised for would have happened without the actor's intentional involvement. (Nolan 2011: 15–17).

I will not assess whether Nolan is correct in all of his claims or whether his classification is correct or not, since it is not crucial for the task at hand in this thesis. What is of interest is Nolan's treatment of the uses 6) and 7). Many writers who link counterfactuals to the issues of causation, causality, causal reasoning or causal explanation (Fogel 1964; Hawthorn 1991, cp. 2; Fearon 1991; Tetlock & Belkin 1996; Stanford 1998: 85–93; Bulhof 1999; Lebow 2007; Megill 2007: 100, 155; Kaye 2010), argue for their position in a way that seems quite loose from the point of view of someone who seriously tries to see whether the proponents are right in their claims. Nolan's arguments are no different in this respect, and can therefore be used to point of the general need for more substantial argumentation on the relation between causal claims and counterfactual claims.

2.3.2 Mysterious link between causal explanatory claims and counterfactual claims

When Nolan discusses what he calls "causation" and causal explanations his aim is to conclude that there are reasons to suggest that there is a connection between causal reasoning and counterfactuals. In the end, however, Nolan does not seem to find good

reasons for this conclusion, and therefore ends up occupying quite vague stance on the question. First, Nolan mentions that it is controversial among philosophers what the connection between causal claims and counterfactuals precisely is (Nolan 2011: 10). Then he notes that "there are arguably several other ways to have causation without this sort of direct counterfactual dependence" (ibid.), but this claim is not presented with any specific examples or argumentation. The only philosophical theory of causation that Nolan mentions is David Lewis's account on counterfactuals and causation, but the conclusion is gloomy: due to the several problems that the account has been shown to face counterfactual analyses are not very popular among philosophers these days (Nolan 2011: 10–11). In spite of these problems that Nolan presents, he still wants to conclude: "the close links between causal and counterfactual judgments mean that historians interested in causation should not reject a role for counterfactuals in historical reasoning" (Nolan 2011: 13). The conclusion is that there seems to be a connection between counterfactuals and causation, but the reasons for why there is such a connection are not spelled out.

The same kind of looseness is apparent when Nolan discusses the role of counterfactuals in causal explanation. Nolan mentions James Woodward's (2003) view that explanations are answers to counterfactual *what-if-things-had-been-different* questions, but he does not pursue the idea any further (Nolan 2011:14). Instead, Nolan continues by claiming that although counterfactual theories of causal explanation might be mistaken (ibid.), "explanations which specify conditions without which the phenomenon to be explained would not have happened seem to be of particular interest" (Nolan 2011: 15). Counterfactuals are deemed as important, but the justification for why they are important is based on a few examples. For the one who doubts the methodological soundness of counterfactual reasoning this might seem far from convincing once again.

My purpose is not to claim that Nolan fails in his argument for the indispensability of counterfactuals altogether, because he is not shedding much light on the connection between causal and explanatory claims and counterfactuals. Nolan's purpose is to show that counterfactuals are ubiquitous in historiography and that there are reasons to think that counterfactuals are methodologically useful in various cases. In this he succeeds. My point is to exemplify the lack of theoretical justification that one comes across when reading through the protagonists' arguments. The idea that there is a

link between counterfactual reasoning and causal reasoning is often brought up, but the idea is not extensively argued for.

In general, it seems that quite often the argument for the connection rests on nothing more but on our everyday tendency to think of causation in counterfactual terms. In addition, in cases where more explicit theoretical ideas of the relation between counterfactuals and causal claims are presented, they are either briefly discussed or just mentioned by referring to unnamed philosophers who think that the counterfactual analysis of causal claims is the right way to go (e.g. Megill 2007: 100). The protagonists are, hence, in need for a philosophical theory of counterfactuals that would turn their idea of the connection between counterfactuals and causation into a serious argument that can be used to defend their position.

2.3.3 Causation, causal relations and causal explanation

The arguments of the protagonists of counterfactuals contain also another possible source of confusion: the terms *causation, causality, causal claims, causal reasoning/thinking* and *(causal) explanations* are used in different ways in different contexts. For instance, counterfactual thinking is sometimes claimed to be relevant for causation whereas sometimes it is presented as important for causal explanation. Sometimes these terms are used interchangeably without paying any special notice to this, whereas sometimes the issues of causation and explanation are discussed separately like in Nolan's paper. Based on the way that these concepts are used in philosophy, the separation is well justified. The terms *causation* and *causality* are usually used to denote the *causal relations in the world,* and the term *causation* is especially used in the context of metaphysics when discussing the nature of these relations.

The true nature of causal relations and the issues of causation are not, however, the issues that concern (or should concern) historians the most. Historians' interest lies in how to answer questions 'what caused *x*?' and 'why *x*?' and the claims '*y* caused *x*' and '*y* explains *x*' that are used to answer these questions. The point of the questions is to find the factors that account for *x*. The point of the answers is to present at least one factor *y* that caused and therefore explains *x*. Historians are therefore interested in i) the specific causal relation between *y* and *x*, ii) the causal and explanatory claims that represent the relation and iii) the explanatory and causal reasoning that is involved in the construction and justification of these claims. The question of counterfactuals

therefore rises up in this realm of causal explanation: what is the relation between counterfactual reasoning and causal explanatory reasoning? Is counterfactual reasoning merely an apt way to approach causal claims and causal reasoning in some cases, or is there actually a conceptual relation between causal claims and counterfactual claims that automatically links all causal reasoning to counterfactual reasoning?

In order to avoid conceptual confusions, I will adopt the following way of using causal terms in this thesis; *causation* and *causality* are terms used to refer to philosophical views on the nature of causal relations. The term *causal relation* is used when speaking of specific causal relations that historians suggest in their *explanatory* and *causal claims* of the general form '*y* explains *x*'. The task at hand is therefore to present a *theory of causal explanation* – not of causation or causality – that can serve to further elaborate the link between causal explanatory claims and counterfactual claims. In this context, when the word *cause* is used it means an *explanatory cause – a factor* that is used to explain certain event *x*. Now that the terminology is in place, it is time to further clarify the protagonists' argument.

2.3.4 Argument for the use of counterfactuals in historiographical causal explanations

The argument of the protagonists of counterfactuals who claim that counterfactuals are methodologically justified in the context of causal explanations can be reconstructed as follows:

1) *The task of the historian*: Historians do not merely describe or recount past events, they also try to account for or explain them.
2) *The nature of explanation in history*: Since explanation in history is causal explanation and the causal explanation of events is the business of identifying causes of events, historians are in the business of making and justifying causal explanatory claims.
3) *The nature of causal claims*: Causal explanatory claims have an inherent conceptual connection to counterfactual claims and therefore every causal claim carriers with it a counterfactual commitment.

The antagonists and protagonists of counterfactuals do not necessarily disagree on the first and second premise. E.H. Carr, despite of his despise for counterfactuals, thought

that it is the task of historian to explain events by specifying their causes (Carr 1964: 105). So, the disagreements most likely concern the third premise: the antagonists and protagonists disagree on the conceptual issue whether causal claims are inherently related to counterfactual claims. As already pointed out, this is the premise that the protagonists of counterfactuals have not paid too much attention to, although it is central to their main claim that counterfactuals are indispensable for a historian.

The third premise says that there is a *conceptual relation* between counterfactual claims and causal explanatory claims. In the context of confirmation of hypotheses this means that once one has justified a causal claim one has justified the related counterfactual as well. This claim does not imply any commitment to a stronger, according to which one can only justify a causal claim by first justifying the counterfactual claim with counterfactual inferences. Although the volumes edited by Tetlock & Belkin (1996) and Tetlock, Lebow and Parker (2006), are dedicated to counterfactual thought experiments that are purported to support specific causal claims, these authors seem not to be committed to the stronger thesis; the are only committed to idea of conceptual relation (Tetlock & Belkin 1996: 3–4; Tetlock & Parker 2006: 17–18).

The conclusion of the chapter is that there is a need for a philosophical theory of causal explanation that explicitly shows how causal claims are related to counterfactual claims. This theory will have an explicative justificatory role from the point of view of the protagonists of counterfactuals. The purpose of this thesis is, however, to go beyond just this mere explicative role and to provide an account of causal explanation and counterfactuals that provides some methodological fruits as well. Showing that – and explaining why – one can methodologically gain something by using counterfactuals further justifies the use of counterfactuals. The following chapters of the thesis are will show that the contrastive counterfactual theory of causal explanation is the theory that fulfills these desiderata.

3 Contrastive counterfactual theory of causal explanation

Counterfactuals, causes, and explanations are three sides of the same strange three-sided coin; you cannot have one without the other two.

– Johannes Bulhof (1999: 147)

The philosophical theory of causal explanations and causal claims I use to pin down the role that counterfactuals play in historiographic explanations can be called *counterfactual theory of causal explanation* (Woodward 2003) or *contrastive counterfactual model* or *account of explanation* (Ylikoski 2001: 7; 2007). Although Woodward's and Ylikoski's accounts of causal explanation are not similar in every detail[1], the accounts share sufficiently similar core ideas. Thus, I see myself entitled to frame them as proponents of the same type of theory, and to use the ideas and arguments of both of them for the purposes of developing my own argument. The contrastive counterfactual theory combines the idea of contrastive *explanandum* with some of the insights that have emerged in the attempts to analyze causal claims and causation[2] in counterfactual terms. The purpose of this chapter is to present the main theses of the account. First, I will briefly present the background of the counterfactual contrastive theory, and the ways in which the ideas central ideas of this theory have been put to use in philosophy of historiography in recent decades. Then I will present the main theses of the account in a more systematic manner.

[1] For instance, Woodward's theory is extremely rich in technical definitions and thoroughness while Ylikoski's account is more accessible in its argumentative clarity. Woodward is on the quest of establishing a theory of *causation* while Ylikoski's interest lies in the establishing a theory of causal explanation that is not depended on what kind of theory of causation we adopt or end up adopting – be it counterfactual, probabilistic, unificationist or something else (Ylikoski 2001: 11). In addition, Ylikoski (2001: 8) applies his view to singular causal explanations, whereas Woodward aims at providing a general theory of causal explanation that applies both to generalizations and to singular facts as explananda (Woodward 2003: 17).

[2] One can distinguish between analyses of causal explanatory claims and analyses of causation as a relation in the world as Ylikoski (2001: 11) does. Analyses of causal explanation are accounts on how causal explanation works and what kind of features causal explanations have, whereas the analyses of causation deal with the metaphysical question on the nature of the causal relations in the world. My focus is on how causal and explanatory inference works in historiography. As to the question of causation, I will adopt Woodward's (2003: ch.6) idea of invariance as a key feature of causal relations, which suffices for the purposes of this thesis.

3.1 Two paths to the contrastive counterfactual theory

3.1.1 Counterfactual theories of causal explanation

The point that causal claims are cashed out in counterfactual terms is not a novel one. As we saw in the previous chapter, the defenders of counterfactual historiography have a strong more or less intuitive idea that causal claims and counterfactual claims hang together in some way. Counterfactual interpretations of causal claims have been around as long as causation and causal claims have been analyzed. David Hume, who has not usually been regarded as a strong defender of counterfactual theories of causal claims, does not seem to have been able to resist the quite natural inclination to interpret causal claims with the help of counterfactuals: "We may define a cause to be *an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second*. Or, in other words, *where, if the first object had not been, the second never had existed*." (Hume 1748: Section VII.)

As already pointed out in the previous chapter, David Lewis (1973a; 1973b; 1986b; 2000) has been the main proponent of the counterfactual theory of causation and causal explanation in recent decades[3]. Despite Lewis's influence in the field of metaphysics of causation, his theory has not been put much to use in philosophy of science. In philosophy of historiography Lewis's views have been only briefly referred to in the contexts defending the use of counterfactuals in causal explanations (e.g: Bulhof 1999: 167; Climo & Howells 1976; Fearon 1991: 176; Tetlock & Belkin 1996: 18, Nolan 2011: 10). The most extensive discussion and criticism of Lewis in the context of historiography can be found in Hawthorn's *Plausible Worlds* (1991: 21–25) and in Elster (1978: ch. 6).

The quite minor impact that Lewis's theory has had can be contrasted with Woodward's impact in philosophy of special sciences in recent years. The fact that Woodward's theory seems to solve most of the conceptual problems that have bugged Lewis's account is a merit in itself and of high interest to a philosopher[4], but much of the charm of Woodward's view is derived from its close connections to work done in

[3] Also Horgan (1989) and Ruben (1994) have proposed versions of counterfactual theory of causal explanation in recent decades.

[4] In fact, in his paper on "Causation as Influence" (2000), Lewis presents the new version of his counterfactual theory in order to avoid the problems that have haunted his earlier account. This new version comes rather close to the main ideas of the contrastive counterfactual theory. In this paper, Lewis defines that causation in terms of "how-how" counterfactual dependence: if *a* had happened *in this way*, then *b* would have happened *in that way*. In other words, the ways in which *b* would happen are systematically dependent on the ways in which *a* happens.

statistics and econometrics. Woodward employs the means that, for instance, econometricians have used to distinguish causal relations from mere correlations and gives these ideas a more general and a more philosophical form. In this way, Woodward manages to present a theory that seems to be applicable way beyond the field of econometrics[5].

Besides the general ideas presented by Raatikainen (2011), there have not been any systematic attempts to see whether Woodward is able to deliver the goods for a historian. Thus, this thesis works as a test bed for the contrastive counterfactual theory in historiography. If the test succeeds, there is a case to claim that the explanatory inferences of historians do not substantially differ from the explanatory practices of other fields although there are surely other kinds of questions and methodological challenges that are more unique to historiography. If the test fails, the antagonists of counterfactuals have a new ace on their sleeves: counterfactuals do not seem to play a significant role in historians' explanatory and causal reasoning.

3.1.2 Contrastive approach

The idea of *contrastive explanandum* has gained popularity in philosophy of science in recent decades. According to the proponents of the contrastive view of explanation, explanations are answers to questions of the form of 'why did *f* occur instead of *c*' (see. e.g. Garfinkel 1981: 21-48; Lipton1990; Ylikoski 2001: 8). The term *f* stands for any property, event, action or any other thing the occurrence of which one is interested in explaining. The contrast of *f* is *c* – a property, event or action that cannot occur together with *f*. In other words, *f* and *c* are mutually exclusive contrasts.

The contrastive account argues that explanation is not about plain facts but about contrastive facts of the type '*f* rather than[6] *c*'. Alan Garfinkel (1981: 21–22) introduces the old joke about the bank robber Willie Sutton as an example of contrastive *explanandum*. A priest went to see Willie Sutton when Willie served time in jail. The priest inquired politely why Willie robbed banks. Willie answered bluntly: "Well, that's where the money is". The joke is based on the different contrasts that the priest's question and Willie's answer have. The priest's question seeks an answer to the contrast

[5] For instance, Woodward's theory has been applied to ecology (Raerinne 2011), equilibrium explanations in economics and ecology (Kuorikoski 2007). Woodward's (2003) own examples taken from the sciences cover e.g. mechanics (12–14), genetics (302–307), and medicine (312–313).

[6] There are several ways to state the contrast of an explanation-seeking question in different languages (van Fraassen 1980; Garfinkel 1981: 25, 29; Ylikoski 2001: 24). For instance, in English one can ask: "Why *f* rather than *c*?", "Why *f* and not *c*?", "Why *f* instead of *c*?".

"why do you rob banks instead of being honest and law-abinding?", whereas Willie answers the question "why to rob banks instead of grocery stores or pharmacies?".

As the above example shows, the advantage of using an explicit contrastive *explanandum* is that a contrastive *explanandum* clarifies what the explanatory-seeking question is about. When the explanandum is stated in contrastive terms, it is also easy to judge whether explanations that seem to be competing actually are competing or not. If explanations answer the same contrastive question, they are competing, but if they give answers to slightly differing contrastive questions they are complementing rather than competing explanations (Ylikoski 2001: 22).

In the context of philosophy of historiography, the gains of using contrastive *explanandum* have not been left unnoticed. Jerouen van Bouvel and Erik Weber (2008) have used the contrastive approach to defend explanatory pluralism[7]. Raymond Martin (1989: 59–65), James Fearon (1991: 172; 1996: 56–59), Aviezer Tucker (1998; 2004: 192–194), Tim de Mey and Eric Weber (2003) all introduce the idea of contrastive *explanandum* in order to tackle various problems related to philosophy of historiography[8]. None of these writers, however, links the idea of contrastive *explanandum* to a counterfactual theory of causal explanation in the way I will present next.

3.2 Main ideas of contrastive counterfactual account

The core theses of the contrastive counterfactual theory of causal explanation can be characterized as follows:

1) Explanations are *answers to questions*. The questions can be formulated in several ways: they can e.g. be *why*-questions or mechanistic *how*-questions (Ylikoski 2001: 21).
2) All adequate causal explanations have *a counterfactual form* (Woodward 1984: 236–245; Woodward 2003: ch. 5; Ylikoski 2001: 8).

[7] The pluralist position maintains that there are no a priori reasons to prefer one type of explanation – be it intentional, structural, "micro", "macro", functional etc. – to any other type of explanation. According to van Bouvel and Weber (2008), the contrastive *explanandum* helps to define what type of explanation might work best in answering the given contrastive explanatory question.

[8] Martin tries to defend the view that weighted explanations in historiography can be objective; Fearon takes note that the use of contrastive *explanandum* can facilitate hypothesis testing in international relations, and that failures in specifying the contrast explicitly can, on the other hand, lead to problems in testing; Tucker tries to show that historians who disagree with each other are usually trying to answer different explanatory questions; de May and Weber tackle the question of causal weighting by using contrastive *explananda* and counterfactuals.

3) All *explananda* are best conceptualized as *contrastive* (Woodward 1984: 245 – 249; Woodward 2003: 145–146, 212–213; Ylikoski 2001: 8, 31).
4) Explanations track *objective relations of dependency* in the world (Woodward 2003: 118–121; Ylikoski 2001: 19). These objective relations of dependency can be characterized as *invariant under some interventions* (Woodward 2003: ch. 6).

My intention is not to argue extensively for these theses since it is done in a detailed way both by Woodward and Ylikoski. The first thesis spells out the main idea of erotetic approach developed by van Fraassen (1980), Tuomela (1980), Achinstein (1983) and Sintonen (1984): explanations are answers to questions and the research proceeds by answering and elaborating questions. What the contrastive counterfactual account adds to this is the idea that explanatory reasoning works with contrastive explanation-seeking questions. We try to find those factors that account for the difference stated in the contrast. In this we engage in counterfactual reasoning: If these factors had been different in some respect(s), the contrast *c* would have occurred instead of the fact *f* that actually occurred.

One consequence of the contrastive idea is the claim that when we are explaining a contrast we are explaining some aspects of a phenomenon or an event or a series of events. Therefore we are not seeking to explain the whole phenomenon, event or series of events (Ylikoski 2001: 20). This means that we can, for instance, explain the occurrence or the non-occurrence of an event or particular properties of it, but not all the properties at once (Woodward 1984: 247–249). This is because the contrast always picks one aspect to be explained: we can ask why the ball is blue instead of pink or then we can ask why the ball is colored instead of it being colorless. These are different questions concerning different aspects of the ball, and therefore, they presumably have different kinds of answers.

The contrastive idea is not a generalized descriptive statement of all the explanatory seeking questions in the world. No one is claiming that all questions *in fact* have an explicit contrastive form. We do not usually state the contrast of our explanation seeking questions explicitly, because the contrast of our question is often obvious in the context where the question is presented. What the contrastive idea is claiming is that all explanation seeking question have an *implicit* contrastive structure that can be spelled out explicitly when needed, for instance, in cases where there is confusion about what purported explanations actually can explain.

3.2.1 Contrastive counterfactual account of singular causal explanation

The contrastive counterfactual theory of causal explanation can be applied both to type causal claims and to singular causal claims. Type causal claims say in generic terms how certain factors of type A (e.g. smoking) cause changes in other factors of type B (e.g. chronic bronchitis). Singular causal explanations, on the other hand, involve token causal claims that state the causes of some particular event, action or series of events. Since historians rarely explain generic repetitive patterns, repetitive phenomena or generalizations, I will limit my focus here on the characterization of the role of counterfactuals in *singular causal explanation* and the analysis of token causal explanatory claims. The central task is to shed light on the ways in which counterfactual reasoning works when answering questions like 'why did this certain event occur?'.

The contrastive counterfactual theory analyzes singular or token causal claims of the form '*a caused b*' as follows:

(3.1) *a* [*a**] caused *b* [*b**]
or interchangeably[9]
(3.2) *a* [*a**] explains *b* [*b**],

In this formulation, *a* is a particular event or action or a set of events or actions [$a_1 \ldots a_n$] that is claimed to have caused, and therefore to explain, a certain particular event or action *b*, whereas *a** and *b** are the contrasts of *a* and *b*. In Woodward's terms, *a* and *b* are the *actual values* of two *variables* A and B. *a** and *b**, on the other hand are alternative values to *a* and *b* that the variables A and B could have taken, but in the actual case did not take[10]. *Variables* are properties (or magnitudes) that can take more than one value (Woodward 2003: 39). For instance we can have the variable of color C that can have several different values [c_1 = red, c_2 = blue, c_3 = orange, c_4 = pink etc.]. The advantage of using the notion 'variable' also in singular causal explanation is the fact that it is an apt way to refer to all the various terms like 'event', 'action', 'process',

[9] As I said in the chapter 2 (section 2.2.3) I use the term *cause* to refer to *explanatory causes*. That is why the verbs *to cause* and *to explain* are used interchangebly.
[10] I will mark variables with capital letters such as A, B, X, Y and the values that these variables take with *a, b, x, y* and the contrastive values with *a**, *b, x**, *y**.

'property' or 'phenomenon' that one uses in one's explanation. I will also speak of *a* and *a** as *facts*[11].

The main crux of this analysis is that singular causal explanations ask why a certain variable took the value *a* it actually did instead of some other value it could have had (why this fact *a* instead of its contrast *a**?). For instance we can ask why the ball is blue instead of being pink. In addition, we could also ask why the ball has a color blue rather than it being colorless, in order to inquire why a certain variable had a value in the first place. The point is that specifying the contrast(s) of the question "why is this ball blue?" helps to clarify what we are in fact asking: are we interested in knowing why the ball has just the color it has or rather why it has a color in the first place?

3.2.2 From causal claims to counterfactuals

The idea of contrastive *explanandum* has been widely promoted in the literature of causal explanation. However, it has not been too common, to state the *explanans* or the token cause in contrastive terms *a*[*a**] as I have done here following Jonathan Schaffer (2005) and Petri Ylikoski (2011). The reason why I have now chosen to present the singular causal claims with double contrasts, is the fact that the formulation shows explicitly the structure of the counterfactual claims that are implied by token causal claims. Put in more exact terms, the claim '*a*[*a**] explains *b* [*b**]' implies the following two counterfactuals:

(3.3) if *a* had not happened, *b* (defined as a contrast to *b**) would not have happened

(3.4) if *a** had happened, *b** would have happened

For instance, in the case of the contrastive question why the ball is blue rather than pink, (which was the original color of the ball), we can say that it is blue because Mary painted it blue. The contrasts of the *explanandum* are [*b* = the ball is blue; *b** = the ball is pink] and the contrasts of the cause that explains the asked contrast are [*a* = Mary painted it blue; *a** = Mary did not paint it at all]. The token causal claim implies two

[11] When a variable A takes a certain value *a* or *a**, this constitutes a fact that *a* or a fact that *a**.

counterfactual claims. 1) If Mary had not painted the ball blue, the ball would not be blue. 2) If Mary had not painted the ball blue, the ball would be pink.

The main plot of the contrastive counterfactual theory runs therefore as follows: in singular causal explanation, we explain the contrast of the *explanandum* by specifying the factor or group of factors that account for the asked difference specified in the contrast. The causal claim states that there is a causal dependency relation between the *explanans* and *explanandum* and the dependency relation can be characterized in counterfactual terms. The formulations (3.3) and (3.4) are simple ways to introduce the basic idea of the counterfactual analysis of causal claims, but they do convey the idea of *intervention* that is central for making the counterfactual analysis to work. It is therefore time to further characterizing the semantic workings of counterfactual claims and to discuss the concept of intervention and the more precise formulation of *dependency relation*.

3.2.3 Tracking invariant dependency relations with interventions

Following Woodward's (2003) *manipulationist account* of causation the truth conditions for singular causal and explanatory claims are characterized as follows:

(3.5) *a*[*a**] causes *b*[*b**] if and only if there is an intervention on the variable A such that: if the intervention changed *a* to *a**, this would bring about a change from *b* to *b**.

What does it mean to intervene on *a*[*a**]? An *intervention* on some value *x* of the *explanans* variable X with respect to the *explanandum* variable Y is *a causal process* that changes the value of X in an appropriately exogenous way, so that if a change in the value of Y occurs, it only occurs in the virtue of the change in the value of X (Woodward 2003: 94). Interventions are also called *ideally surgical manipulations* (e.g, Kuorikoski 2010: 23) that affect only the *explanans* variable and leave the values of all other variables intact and show what kind of effect the manipulation of the *explanans* variable has on the *explanandum* variable.

Since interventions are characterized with the help of other causal concepts, the characterization (3.5) cannot provide a reductive analysis of the concept of cause. Another interesting point is what the characterization (3.5) suggests about "the origins" of our causal reasoning: the way in which we humans infer and reason about causal

relations has its roots in our ability to act as manipulating agents who actively intervene on the workings of the world. We want knowledge of the world that helps us to manipulate our surroundings, that is, causal knowledge. Our beliefs and ways of reasoning, however, do not determine the truth of the causal claims that are of interest to us: it does not depend on the beliefs and reasoning of the agents whether bringing about *a*[*a**] brings about *b*[*b**]. The condition of a successful manipulation is therefore that there exists a causal dependency relation that our causal claims and the related counterfactuals manage to capture. Due to this, the contrastive counterfactual theory is committed to a realist stance (Woodward 2003: 118–119).

Another important point of Woodward's manipulationist account is that its applicability is not limited to those situations in which we are actually in a position to manipulate certain variables. This is good news for the disciplines such as historiography, various social sciences, ecology, or cosmology whose subject matters hardly provide any opportunity for straightforward experimental manipulations. The wide applicability of Woodward's account stems from his idea of *intervention* that is not in any special way connected to human action. An intervention can be *any causal process* that *changes* the value of X and controls the values of other variables $Z_1 \ldots Z_2$ so that X becomes the only variable that the values of Y are dependent on. Thus, events of nature that occur without any human involvement as well as experimental manipulations can both be regarded as interventions if they work in the way that interventions should: they turn the variable X into "a switch" that controls the values taken by the variable Y.

The idea of intervention also gives the criteria to distinguish real causal relations from correlations and other relations that only appear to be causal: mere correlations cannot be used to successfully manipulate *b*[*b**]. Woodward (2003: ch. 6) calls the property that makes successful manipulation possible *invariance under interventions*. For a relation to pass as invariant there must be at least one intervention on the *explanans* variable that changes the value of the *explanandum* variable in the way that the causal claim says it will. Mere correlations do not pass this test, whereas real causal relations do. In this way Woodward's account solves one of the key problems that has haunted several other analyses of causation and causal explanation.

The well-known example of a barometer reading predicting the weather but not explaining it serves as an example that nicely illustrates the idea of intervention. We can manipulate the values of the barometer but not one of these manipulations seems to

have any systematic effect on the weather – changing the barometer reading from 1040 mbar to 950 mbar does not make fair weather turn into a storm and one should not expect a nice weather for an outdoor day after tempering with the readings of one's barometer. The relation between the barometer readings and the weather shows no invariance. There are no interventions on the barometer readings that would change the weather. The relation is a mere correlation.

Instead, the relation between air pressure and weather exhibits a certain degree of invariance. There are *interventions on the air pressure* that if they *were carried out* in some way they *would change* the weather. For instance, if there was an intervention that would change the high pressure to extreme low pressure, there would be a change from fair weather to rain or storm. So, the variable whose changes explain the contrast between fair weather and rainy weather is the air pressure, not the barometer readings for which the air pressure also accounts.

The above example also helps to illustrate the feature that distinguishes Woodward's account from standard naïve analyses of causation in counterfactual terms: the account saves the asymmetry of causal relations and does not allow backtracking counterfactuals. The naïve counterfactual reasoning that follows the lines of the characterizations (3.3) and (3.4) easily leads to results where the cause could be deemed as counterfactually dependent on the effect. In other words, if we took counterfactuals like (3.3) and (3.4) to be sufficient to determine causal relations, we would have causal relations where the effect would be the cause of the cause! For instance, we could infer that barometer readings are the cause of the low pressure with the following inference: if the barometer readings had been low, there would certainly have been a low pressure.

This example brings out the well-known point that not all counterfactual dependencies correspond to causal or explanatory relationships. Epistemically, it is justifiable to infer the occurrence of storms based on the barometer readings, but these inferences do not reflect causal relations in the world. Woodward's notion of intervention is, therefore, used to distinguish the true causal relations and the related counterfactuals from the counterfactual inferences that do not correspond to causal relations.

As already mentioned, in the case of singular causal claims and singular causal explanations causal explanatory claims can be characterized in interventionist terms as follows:

> *a*[*a**] causes *b*[*b**] if and only if there is an intervention on the variable A such that: if the intervention changed *a* to *a**, this would bring about a change from *b* to *b**.

This condition clause states that there should be an invariant relation between *a*[*a**] and *b*[*b**] if the relation is to be counted as causal. Thus, in order for a causal claim to be explanatory it should support the right sort of counterfactuals about what would happen under interventions. So, Woodward's manipulationist account does not merely say that causal claims should support any counterfactuals but rather that they should support *counterfactuals formulated in interventionist terms.* I call these counterfactuals that are needed in causal explanation *interventionist counterfactuals*. By making use of interventionist counterfactuals we can track invariant dependence relations in the world, which are, the truth-makers of causal explanatory claims. Interventions and invariant dependency relations will be more thoroughly discussed in chapter 5 that will focus on the epistemology and methodology of counterfactual inferences in historiography. Next, however, we will move on to examine how the general framework and conceptual tools presented in this chapter solve some of the central problems of historiographical explanation.

4 Contrastive counterfactual theory in historiography

'If there is such a thing as a sense of reality', Robert Musil had agreed in The Man Without Qualities, … '– then there must also be something that one can call the sense of possibility'. But in history and social sciences it has not been usually clear what there is to be gained in developing it.

– Geoffrey Hawthorn (1991: 4)

When browsing through the literature on explanation in historiography, one can't help getting the impression that there are as many philosophies and methodologies of historiography as there are philosophers and historians writing on them. The purpose of this chapter is to lay the basis on a coherent and unified view of causal explanation in historiography based on the contrastive counterfactual theory introduced in the previous chapter. Various themes ranging from the problems of explanatory selection, causal background and negative causes to the problem objectivity and subjectivity are discussed in order to show how they can all be fruitfully handled with the idea of contrastive *explanandum* and contrastive causal claims.

In the end of the chapter, the contrastive counterfactual model is compared to the best known, although conflicting, philosophical accounts of historiographical explanation: the covering law model and the non-causalist or intentionalist view of explaining actions. The covering law model is shown to face some serious problems, and therefore, not to be expected to provide fruitful explanatory methodology for historiographical explanation in the same way that the contrastive counterfactual model provides. The non-causalist position is answered by arguing that reasons can be treated as causes in the framework of the contrastive counterfactual theory without undermining their nature as reasons.

The overall argument of the chapter is the following: on the basis of the discussions in section 4.1 and on the basis of the comparisons to the best-known alternative theories of historiographical explanation in section 4.2 the contrastive counterfactual theory can be regarded as the most fruitful theory to characterize the nature and problems of historiographical explanation.

4.1 Contrastive *explananda* and contrastive causal claims in historiography

4.1.1 The contrastive explanation of facts

It is necessary to start with specifying what historians are explaining in their explanations. The general term I use to denote the target of explanation is *fact.* Facts are events, series of events, processes or phenomena that historians are interested in[1]. In woodwardian terms facts are constituted by variables taking a certain values. I will call historiographical claims that refer to facts that have happened in the past, *historiographical descriptions*. As already pointed out in chapter 2, historiographical descriptions are claims of the facts of the past and they are the end result of historiographical research; historians infer what happened in the past with the help of the traces of that past salvaged to this day (Tucker 2004: 17, Collingwood 1994[1946]: 251–252).

As also pointed out in chapter 2, an interesting feature of historiographical study is that the problem of underdetermination is often faced also in practice due to lack of evidence. The lack of evidence is one reason why historians often disagree on purported facts without reaching an agreement. For instance, the gaps in data make it a detective story to find out whether there were peasant revolts in the Classic Period Lowland Maya civilization as J.E.S Thompson (1966) suggests. Some facts have, of course, strong evidential support and most historians agree on them. For instance, the fact that John F. Kennedy became the 35th President of the United States on January 20th 1961 is hardly under dispute.

In cases where historians have reached an agreement on what happened, they might still argue on what *to call* the things that happened. Indeed, the most famous historiographical disputes have centered on the names given to the long term processes such as “the First Industrial Revolution” or “the fall of the Roman Empire”. In these cases, the debate has mainly been on whether the terms “revolution” or “fall” capture the nature of the changes that are inferred from the data. One task of philosophy of historiography would then be to explore how these kind of descriptive disputes develop

[1] There is a general discussion within philosophy of science whether facts should be considered as the right relata for causal explanation (see e.g. Mackie 1974: ch. 10 and Ruben 1990: ch. 5). In the metaphysics of causation, the question has been whether the relata of causation are facts (see e.g Mellor 1995). I do not have the opportunity to pursue these questions any further in this thesis where the central question is causal explanation in historiography, not causal explanation in general.

and how they are linked to explanatory competition[2]. Unfortunately, in order to keep my lines of argument together, I do not have the possibility to explore this question here. Another important issue I cannot discuss in detail is the question how counterfactual and causal reasoning might work together when assessing whether the pieces of evidence, which one has, provide reliable evidence for a certain fact or not (see section 2.2.3 however). Thus, the explanations that are the focus of my thesis are *the explanations of the facts that historians have inferred from the data and which they mainly agree on.*

The starting point in applying the contrastive counterfactual theory to historiography is the idea that historians explain facts - events, series of events, processes or phenomena of the past – and that these facts should be conceptualized in contrastive terms[3]. Take, for instance, the presidency of John F. Kennedy. We can ask why John. F. Kennedy won the Presidential election of 1960, and implicitly assume the following contrastive question: why did John. F. Kennedy win the Presidential election of 1960 instead of Richard Nixon? The explanatory task for the political historian would then be to try to pick out the factors that caused the fact that it was JFK who gave the inaugural address on January 20th 1961 instead of Nixon. This explanatory information would ideally be of the sort that it would allow the historian to answer *what if things had been different*-questions such as "what if Kennedy had performed as badly as Nixon did in the first televised U.S. presidential debates? Would he have been elected then?". This example shows how the idea of contrastive *explanandum* leads to counterfactual considerations.

My point here is that it is not just the cases where the contrast seems fairly obvious that should be interpreted in contrastive terms. For instance, the classical question "Why did the Greeks win the Battle of Salamis in 480 BCE?" can be thought to have several contrasts. The most obvious contrast in this case is: "Why did the Greeks win the Battle of Salamis instead of the Persians?". But there are also other contrasts that intrigue the historians such as: "Why did the Greeks win the Battle of Salamis and not the previous sea battles?" or "Why did the Greeks win the Battle of Salamis quite overwhelmingly and not just barely?". It is interesting to note that a

[2] Raymond Martin (1989: ch.3) has taken a step into this direction with the case study on the disputes of archeologists and historians on the population loss of the Classic Period Lowland Maya civilization

[3] The natural idea is that explanations cannot explain the things of the past in themselves. The past events need be conceptualized in some way to be a target of explanation (see Ylikoski 2001: 20). The idea of contrastive *explanadum* is a way to conceptualize facts for explanatory purposes.

simple-looking question like “Why did the Greeks win the Battle of Salamis?” can have various contrasts, and due to this different meanings, but the point goes deeper here. What is worth noticing is the fact that these different contrastive questions usually require somewhat different kinds of answers. The contrast thus determines what kind of factors one should include in one’s explanation of the contrast – and it might quite often be that these are not the same factors in all the contrasts.

There are numerous examples of contrastive questions; indeed, if the hypothesis of the proponents of the contrastive approach is correct there are at least as many examples than there are causal explanatory-seeking questions. This is why I will settle for one more: the process called the First Industrial Revolution. The point of this example is to make clear “the corollary” of the contrastive idea – the contrastive idea implies that we are explaining aspects of events or series of events, not the events or series of events as a whole. The reader might have noticed this with the previous examples, but with a long-time process consisting of thousands of events like the First Industrial Revolution one cannot expect to be able to provide an all-compassing narrative of the things that happened and link them causally together as a total complete explanation. What some economic historians have done instead is that they have explained the process of British industrialization through contrastive questions. They have asked: “Why did Britain industrialize instead of other countries in continental Europe?” (Crouzet 2001: 111: 114), “Why did the cotton industry in Britain become factory based instead of India taking the lead?” (Landes 1998: 225–30), and “Why Western Europe and not China?” (Pomeranz 2006).

4.1.2 Explanatory selection in fixing the causal background

What would be wrong in trying to explain an event as a whole? In principle, nothing if we take the explanation of the whole event to mean that we would be able to explain all the contrasts of all the aspects of an event (Ylikoski 2001: 20). In practice this, of course, might turn out to be an insurmountable task. Consider, for instance, what it would mean to explain all the contrasts of all the aspects (like cotton industry, mining, organization of labor, the institutional context) of the First Industrial Revolution. Carl Hempel presents the problem of complete or total explanations nicely as follows:

> What is sometimes called the complete description of an individual event (such as the earthquake of San Francisco in 1906 or the assassination of Julius Caesar) would require a statement of all the properties exhibited by the spatial region or the individual object involved, for the period of time occupied by the event in question. Such a task can never be completely accomplished.
>
> (Hempel 1942: 37.)

Peter Railton's (1980) idea of *ideal explanatory text* is one way of defining what it would mean to explain events as a whole. According to Railton *ideal explanatory text* is a theoretical concept that includes *all accurate information about every due-to relation relevant for the explanandum* (Railton 1980: 147-148). All information that reduces insecurity of the things that account for *explanandum* is deemed as explanatory. Førland (2004) therefore praises Railton's account since it "invites the historian to search for ever more explanatory information". Seeking explanatory information is naturally a good thing, but thinking in Railton's terms leads to the problem of explanatory selection.

The problem of explanatory selection comes from the fact that *we do not deem all causal information as explanatory relevant* (Ylikoski 2001: 20; Hesslow 1983). We can describe a vast array of events and processes leading to an event we are interested in explaining just as Hempel points out. How then to select which of the factors of the causal history to include in our explanation? What factors count as truly explanatory? The attempt to describe the whole causal history would be practically impossible, as Railton himself also acknowledges, but also if one were to know the whole causal story leading to an event, that would still not count as a proper explanation. Thinking in terms of the phrase "more causal knowledge the better for the explanation" might lead to situations of pragmatic explanatory failures (Ylikoski 2001: 26) where one simply gives too much information in answering a certain question. If I am interested in knowing why John F. Kennedy was elected President in 1960, it is no use telling me the whole history of the U.S presidential elections although the events related to the creation of the election system surely belong to the causal processes leading to Kennedy's election.

The problem of explanatory selection is common to all fields of study. A historian who wants to write a traditional narrative of some event of interest and explain the event faces the problem of choosing what parts to include into the narrative and what to leave out (Little 2010: 14–15). As Daniel Little (2010: 30) puts it:

> "[A] crucial and unavoidable feature of narrative history is the fact of selectivity. The narrative historian is forced to make choices and selections at every stage: between 'significant' and 'insignificant', between 'sideshow' and 'main event' and between levels of description. (Is World War II better described at the level of generals and policy-makers or infantrymen and factory workers?)
>
> (Little 2010: 30.)

The contrastive approach offers a simple way of replying to Little's question. If one is interested in explaining some aspects of World War II, using contrastive questions helps one define what factors are causally significant and what are not and what is the right level of description since it depends on the contrast one is asking. In the Kennedy case the information on the creation of the U.S presidential election system is irrelevant for the question why John. F. Kennedy won the presidential election of 1960 instead of Richard Nixon, but it might be relevant for other contrastive questions. If I were keen on knowing why the U.S. has an election system where it is possible for Kennedy to win the elections although he did not win in over 50 percent of the states, the information on the creation of the U.S. presidential election system would be of high interest, since it would explain why it is so that the U.S. has the electoral system it does instead of some other system. The question concerning Kennedy's victory presupposes that there exist some electoral system in the United States, but explaining this presupposition is giving answer to a different question than the question why Kennedy won instead of Nixon.

The contrastive approach therefore implies that we do not need to be able to list all the complex causal processes leading to an event in order to explain a certain contrast. We only want to know those causal factors that account for the difference stated in the contrast. What is the fact *a* that makes the difference between *b* and the contrast *b**? Contrastive explanatory questions are therefore means to nail down *the causal background*[4] of the explanation. By causal background I mean the causal processes and events that one does not need to state or even know in order to explain why *b* rather than *b**. This is because in the contrastive approach the causal background consists of the processes that *b* and *b** share in common. For instance in the Kennedy case the U.S electoral system is the same for the fact "Kennedy won" and its contrast "Nixon won". Therefore there is no need to state the development of the electoral

[4] The causal background is often also called *causal field* following Mackie (1974: 34–35)

system in the explanation, since it does not account for the contrast of the explanation-seeking question. In this case, the features of the electoral system and their development belong to the causal background.

4.1.3 Negative causes and the principled epistemological objection reconsidered

The reader might remember the principled epistemic objection against the use of counterfactuals in historiography presented in section 2.2.3. According to the objection, we cannot know the things that did not happen, since we do not have a causal access to them. The antagonists of counterfactuals consider this to mean that we cannot have any knowledge of the antecedent or consequents of counterfactual claims since, by definition, the antecedents and consequents of counterfactuals are things that did not happen. Thus it is beyond the reach of empirical sciences to arrive at informed conclusions of the truth-values of suggested counterfactuals.

I think that we are now in a position to answer the principled epistemic objection. Counterfactuals that are conceptually linked to causal claims are epistemically respectable, since causal relations that the causal claims refer to are entities in the world that we have an epistemic access to: historiographical evidence can be used to make inferences on the past causal claims and the related counterfactuals. If the change from *a* to *a** brings about the change from *b* to *b** in the variable B, there is a invariant causal relation between *a*[*a**] and *b*[*b**]. There is nothing counter to facts in causal invariant dependency relations themselves (Ylikoski 2011: 165) – they are relations in this world, and one feature of them is that they support interventionist counterfactuals. Therefore, interventionist counterfactuals can also be said to have truth values – another questions is, of course, whether we have enough evidence to infer the truth values of individual counterfactuals.

The above solution also brings clarity to the fact that sciences seem to be full of sensible claims about things that did not happen, and to which we do not have a direct epistemic access, such as negative causes, gaps, omissions, preventions and absences (Ylikoski 2001: 37). With the help of the idea of causal claims with contrastive *explanans* and *explanandum* it is easy to see why negative causes do not prove to be

insurmountable for scientific causal reasoning. Take for instance the following causal claim and the related counterfactual scrutinized by Lebow & Gross Stein (1996)[5]:

(4.1) The fact that Kennedy *did not show* enough resolve prior to the Cuban missile crisis caused Khrushchev to deploy the missiles.

(4.2) Had Kennedy shown more resolve prior to the Cuban missile crisis it *would have prevented* Khrushchev from deploying missiles in Cuba.

What is evident in Lebow's and Gross Stein's study is that they do not concentrate on the verbal expressions such as '*did not show*' or '*would have prevented*'. In fact, they formulate (4.2) without the term 'prevention':

(4.3) Had Kennedy displayed greater resolve prior to the crisis, Khrushchev would not have sent missiles to Cuba (Lebow and Gross Stein 1996:124).

This makes it clear that our worries should not center on words or the exact formulations of sentences, but we should rather focus on the contrastive pairs implied by the causal claim. Thus, from the perspective of contrastive counterfactual theory, Lebow and Gross Stein are interested in the relationship of the contrast of the purported cause variable [*c = Kennedy not showing resolve; c* = Kennedy showing resolve*] to the contrast of the effect variable [*d = Khrushchev deploying missiles; d* = Khrushchev not deploying missiles*]. It is also important to note that in this case the fact *c* can be formulated either as an absence [Kennedy *did not show* enough resolve] or as a positive instance [Kennedy displaying greater resolve][6]. On the basis of the new evidence on Khrushchev's aims and beliefs, Lebow and Gross Stein end up concluding that the counterfactual (4.2) is false. The right counterfactual has the contrasts other way around: it was precisely because Khrushchev belied the United States to be superior, and therefore, to pose a serious threat to the Soviet Union under the Kennedy

[5] Lebow and Gross Stein do not present the causal claim (4.1) themselves, their interest lies in the counterfactual claim (4.2), but as I interpret them, they seem to provide evidence that works for determining the truth of (4.1) as well as of (4.2).

[6] The contrastive counterfactual theory also gives a simple way of representing omissions, negative causes, absences and representations. We can think of a variable A taking the value *a*, for instance $a = 0$, that represents an omission or prevention or any other type of negative cause. The variable itself "exists", it just does not have any positive value in the case of negative causes.

administration that he decided to send the missiles to Cuba (Lebow & Gross Stein 1996: 135).

Lebow and Gross Stein seem to have quite a strong epistemic grip of the counterfactuals they assess. My claim is that this is so because in the end their argument concerns the actual causal relations that took place during the Cuban Missile Crisis. The counterfactuals can be employed in order to study these relations by using the evidence there is on Kennedy's, Khrushchev's and their advisors' actual aims, goals and decisions. The antecedents and consequents of counterfactuals are therefore not out of this world and it is not beyond the reach of our epistemic capabilities to evaluate the truth of suggested counterfactuals.

So, what is to be replied to A.J.P Taylor (1969: 16–17) who worries on counterfactuals echoing the strong epistemological thesis presented in chapter 2: "How can we decide on something that did not happen? Heaven knows, we have difficulty enough in deciding what did happen"? A simple answer is that if there is enough evidence that gives us knowledge on the factual causal relations, we are also gaining counterfactual knowledge. What is more, there are cases as the Lebow's and Gross Stein's study presented here in which starting from counterfactual claims is a fruitful way to approach the causal relations one is interested in. Contrary-to-facts is one way to study the facts.

4.1.4 Objectivity and subjectivity in historiography

One question of importance in historiography has been the issue of subjectivity and objectivity of historiographical research. A source of confusion in this debate has been the fact that the terms 'subjectivity' and 'objectivity' have not been very well defined (Martin 1989: 85–86). It is, indeed, so that the debate on subjectivity and objectivity can concentrate on many different, but in some cases interrelated topics such as the role of values in research and the role of interpretation in historiography. My intention is not to get deeply immersed in these themes but instead to make a few comments on the debate based on the contrastive counterfactual approach.

In the context of historiographical explanations surprisingly many historians have contented that historiography is inescapably subjective, whereas philosophers have been quick to defend the idea that *in principle* historiography can be objective (Martin 1989: 85–86). It is worth separating three issues in this context:

(S1) *Subjectivity in choosing research questions*: There is no other way of deciding what to study and what to seek explanations for than the researcher's interests and values or the interests or values of the research community or society as a whole.

(S2) *Subjectivity in selecting explanatory factors*: the selection of explanatory facts and factors that one includes in one's explanation depends on subjective considerations such as researchers' tastes.

(S3) *Subjectivity in evaluating proposed explanations*: The goodness of one's explanation is determined on subjective grounds. This also implies that it is a matter of subjective taste to prefer one explanation to another.

When it comes to choosing one's explanation-seeking questions, the counterfactual contrastive model fits together with the first account of subjectivity. S*ubjective* or *pragmatic* or *value* considerations can be said to play a role in choosing the topic of one's study and in formulating one's (contrastive) explanation seeking questions.

When it comes to assessing counterfactual claims as causal claims, the contrastive counterfactual model takes a non-subjectivist stance towards (S2) and (S3). First of all, the idea of explanatory selection presented earlier makes the rejection of (S2) quite obvious: *once the contrastive explanation seeking question has been fixed*, the selection of explanatory factors is an objective matter (Ylikoski 2001: 36; Northcott 2008: 80). We should "pick" or select those factors and facts in the explanation that make the difference between the occurred fact and its contrast. This also means objecting to (S3): the goodness of the proposed explanation depends on the question whether it is so that the facts that one has selected in one's explanation account for the asked contrast. When there are two or more competing and mutually exclusive explanations that seek to answer the same contrastive question, there is no reason to base the choice between competing explanations on subjective considerations: the advice is to seek evidence that would support or falsify the explanatory role of the proposed explanatory facts.

E.H. Carr (1964: 96) has dismissed all counterfactual reasoning in historiography as the "might-have-been" school of *emotion*. But, even if a counterfactual claim was derived from emotional reactions or other value-laden considerations, the origins of the claim should not be used to condemn it as unscientific or unhistorical, as Carr (1964: 97) does. When it comes to counterfactual questions like

"what if the Russian revolution had not been?", the reason for presenting this kind of question might derive from the emotional responses of the victims of Bolshevism as Carr (ibid.) suggests. Another matter then is whether this kind of question is to be interpreted as a causal question that seeks to assess seriously the causal importance of the Russian Revolution for some aspects of the events that followed the Revolution (for more elucidation on this, see chapter 6). The contrastive counterfactual theory regards this question as a proper causal question. The alternative is just what Carr (1964: 97) claims: that counterfactual claims and questions are parts of parlour-games – a form of "iffing" that people constantly engage in in their everyday lives for amusement or for comfort without any real interest in the truth of the counterfactuals as causal claims. But not all counterfactual reasoning is of this second type, as Carr believes. Counterfactual reasoning goes hand in hand with causal reasoning, and in this context the justification of the counterfactual claims should and can go beyond subjective iffing, even if the counterfactual claims would stem from clearly subjective interests.

By denying the subjective claims (S2) and (S3) I am now making a *principled* point. This means objecting to the ideas according to which there are conceptual or principled reasons why explanatory selection and evaluation of explanations are inherently subjective in historiography or in all fields of research in general. My intention is not to dismiss in any way the problems one encounters in actual historiographical research, since there might be empirical reasons why the selection of explanatory factors and evaluation of competing explanations can in some cases be extremely difficult. This might be due to the lack of evidence and unsettled disputes on what some specific piece of evidence is evidence for. There are also important epistemological and methodological questions I have not touched yet. How to come to the conclusion that the proposed factors truly account for the difference? On what kind of inferences and evidence one should base one's evaluations of the goodness of one's explanations? These questions will be further discussed in the next chapter.

Raymond Martin (1989: 53–63) defends the objectivity of historiographical explanations in a very similar way as I have done here. Martin's argument is targeted towards William Dray who has claimed that the impossibility of distinguishing causal background conditions from explanatory causes makes explanatory selection a matter of subjective taste. Martin tackles Dray's idea by suggesting:

> "… 'the cause' is selected from among the contributory causes on the basis of comparison between a situation in which the result to be explained occurred and some other situation which I shall call 'the comparison-situation'. … 'The cause' then, like the conditions contributes to the result, but unlike the conditions, also differentiates between the situation in which the result occurred and the comparison-situation."[7]
>
> (Martin 1989: 59.)

The idea of using a comparison situation or a contrast in specifying the explanation-seeking question allows Martin (1989: 62) to conclude: "it is always possible to distinguish causes from conditions on factual grounds by addressing an appropriately formulated explanatory question". Dray's idea that historians are forced to distinguish explanatory causes from conditions on moral or other normative grounds is thus undermined.

There might be many historiographical studies in which the selection of causes has happened on moral or political grounds as Dray suggest to have happened in various interpretations of the American Civil War (1962) or in the debate surrounding A.J.P Taylor's interpretation of Hitler's role in the outbreak of the Second World War (1978). Dray's idea (1978: 150–151) is that Taylor and his critics engage in causal reasoning in the way that lawyers do: they seek to attribute moral responsibility, blame and appraise[8]. Attributing moral responsibility seems to require causal reasoning, but unlike Dray, I see it highly problematic to think that attributions of moral responsibility should guide the causal reasoning of a historian. Following the method of a historical prosecutor leads quite easily to the situation where only causal relations relevant to moral attributions are considered as the only causal relations that matter. For instance, in Taylor's case the question what the causal role that Hitler played in outbreak of the Second World War, has become merely the question of whether or not Hitler *intended* to cause the Second World War. The question of Hitler's intentions is of high relevance when assessing Hitler's responsibility in the start of the Second World War, but there are causal processes leading to the Second World War that Hitler participated in and in

[7] Martin (1989: 159) calls this view "the consensus view" which he bases on ideas of philosopher's such as C.J. Ducasse, J. Feinberg, S. Gorowitz, H.L.A Hart, A. Honeré, M. Scriven, R. Shope, and M. White.

[8] The connexion between causal explanations and attributions of moral responsibility in the context of law and in history is extensively discussed by H.L.A Hart and Tony Honoré (1985). Hart's and Honorés central idea is that historians often think of causal explanation in the same way as lawyers do.

which Hitler's causal role can be determined without knowing his intentions or motives[9].

Therefore, there is a need to distinguish the two roles of a historian: the researcher aiming at causal explanations and the prosecutor aiming at moral convictions. If a historian seriously wants to act as a prosecutor, she should make this aim explicit, bring forth the causal relations of the case she is working on, and then pin down the causal relations that she considers to be relevant for attributions of moral responsibility and the reasons she has for considering just these causal relations relevant for solving questions of moral responsibility. The worst thing that a historian could do is to act as a bad historical judge who interprets the evidence and selects the facts that she deems as explanatory on the basis of *a priori* considerations of who is guilty and who is to be praised. This kind of method is bound to lead to serious subjective biases in historiography. However, as I, as well as Martin, have aimed at showing this is not the method that a historian needs to embrace.

4.2. Facing the competition: non-causalism and covering laws

4.2.1 Intentional explanations and causal explanations

William Dray (1963) has argued, following Collingwood (1994[1946]), that *Geisteswissenschaften*, the sciences of the mind, do not involve causal explanation. According to Dray, the sciences of the mind, such as historiography, focus on the behavior of human beings in so far as it counts as *action* or as an expression of *rational thought* (Dray 1963: 108 ; D'Oro 2011: 144). The rational explanation of action entails constructing an argument – a practical syllogism – that shows that there is a rational conceptual relation between the agent's action and the agent's beliefs and desires/motives/goals. In rational explanation the relation between *explanans* and *explanandum* is therefore not an empirical or causal connection, but a conceptual relation that is established through reflection or understanding (Dray 1963: 108; D'Oro 2011: 144). Other philosophers who have developed a non-causalist position analogous to Dray's and Collingwood's include, to mention a few, Peter Winch (1990[1958]) and

[9] For instance, Hitler's command to attack to Poland had the effect of starting the process of declarations of war that cumulated into the Second World War. In this case, we do not need to know Hitler's actual intentions in order to say that this act of Hitler sparked off the Second World War. For, the assessment of Hitler's moral responsibility it might be of interest to know what Hitler was actually aiming at by giving the order to attack to Poland.

Georg Henrik von Wright (1971). How does this position relate to the contrastive counterfactual theory?

First of all, the contrastive counterfactual account is easily adapted to give a causal interpretation of intentional explanations (Ylikoski 2001: 96–100). Intentional explanations concern practical reasoning where an agent deliberates based on her desires, motives and beliefs what action to take. The contrastive counterfactual theory deems as causally explanatory all factors that make a difference stated in the *explanandum*. If the *explanandum* consists of the question why the agent H did the action *c* instead of *c**, we quite often consider the agent's beliefs and desires to be the factors that account for this difference. Thus, in the case of practical reasoning, the contrastive counterfactual account is in line with our everyday activity of explaining human actions (Ylikoski 2001: 96). We explain our own action as well the actions of others with the counterfactual framework all the time: if she had believed differently, known this or had her motives been different, she would not have acted in the way she did. Therefore, a historian who wants to know why Kennedy did not want to launch an airstrike against the missile centers in Cuba during the Cuban Missile Crisis *instead of* taking the advice of some of his advisors and do so, is causally explaining Kennedy's action when she answers the question by citing Kennedy's beliefs and goals.

It is important to notice that this analysis is compatible with the idea that human action is not the same as the movement of mindless blocks: human action can said to be conceptually connected to ideas of rationality, deliberation and freedom as non-causalists take it to be. However, the contrastive counterfactual account refutes the thesis "reasons are not causes" as an either-or distinction[10]. Reasons can be taken as causes if they account for the contrast in the *explanandum*. By claiming that reasons can be causes, I am not saying that the reasons would stop being reasons in the context of causal explanations. Reasons *qua* causes have similar kinds of counterfactual dependency relations to action than all other types of causes have to their effects. The causal reasoning with reasons works with the same interventionist idea: If there had been an intervention on the agent's reason *r*[*r**], she would have done action *c** instead of *c*. This interventionist interpretation does not imply that the agent's reasons should be subjected to wicked external manipulations performed by other agents. The intervention

[10] Donald Davidson (1963) has extensively argued for this position.

can also be the agent's own change of mind. Thus, the intervention can come from various sources. All that is required is that it makes *r*[*r**] a switch for *c*[*c**].

I simply cannot discuss all the rich arguments that the non-causalists presented to support their position that causal explanation does not coincide with explaining (or understanding) human action. In any case, it seems that the contrastive counterfactual account is compatible with the main idea of the non-causalists: there can be conceptual relations between the *explanans* and *explanandum* but that does not exclude the possibility that there can also be causal relations between them. It can also be that in order to verify the reasons for a certain action we need to have an idea what the action was about and vice-versa, as von Wright (1971: ch. 3) suggests, but this does not mean that there could not be counterfactual dependency relations between reasons qua causes and action[11].

In order to grasp why the non-causalists had reasons to argue for a distinct method for explaining human conduct, one needs to point out that the theory of explanation that they contrasted their views with, was not the one presented here, but the covering law model. It was the covering law model that many non-causalists took to provide the framework for causal explanation[12]. Since the non-causalists regarded the covering law model as an implausible candidate for explaining human behavior for various sorts of reasons, they had a good case to argue for a non-causal theory of explaining human action. So, let us turn to the covering law in order to grasp the theory of scientific explanation that the non-causalists were opposing in the spirit of anti-positivism and German tradition of hermeneutics.

[11] Not to give a wrong impression, there is a need to stress the following: there can be cases where the conceptual link between the *explanans* and *explanandum* is too tight to provide the kind of interventionist counterfactual dependence that is required in causal explanations. For instance, the agent's *intention of greeting* a friend is conceptually too close to be an explanatory cause to *the greeting* the agent gave. This would be tantamount to saying that I greeted because I wanted to greet, which is not very illuminating. (cf. von Wright 1971: 123–124). However, there is no conceptually dubious connection between the intention of catching the friend's attention and greeting the friend. Therefore the contrast [the belief that greeting a friend is an apt way to catch the friend's attention; the belief that greeting a friend is not needed to catch the friend's attention] can be used to causally explain [the greeting; not greeting] without any problems. It is true that the content of the belief conceptually constitutes the act of raising the hand as an intentional action: it was *a greeting*, in contrast to a compulsory movement and other plain physical bodily movements. But since the changes in the belief also bear right kind of counterfactual relations to the *explanandum*, the belief can in this case presented as an explanatory cause.

[12] It might be a mistake to interpret Hempel as providing a model for *causal* explanation, since the covering law model does not operate with or try to explicate the concept of 'cause' at all.

4.2.2 Debunking the covering law model

In 1942, Carl Hempel published the article “The Function of General Laws in History”. It was the first publication on the model of scientific explanation that came to be known as the Hempel-Oppenheim-model, deductive-nomological model or the covering law model. In philosophy of historiography it launched a debate that flourished especially in the 1950’s, 60’s and 70’s, but which still seems not to have ceased. Articles and books were written pro (e.g. White 1965; Murphey 1973; Roberts 1996) and against (e.g. Scriven 1959b: 457; Donaghan 1964; Danto 1965: 211) the applicability of the covering law model in historiography.

The covering law model is meant to provide a universal account of scientific explanation that would apply to all fields: various natural sciences, social sciences, and as explicitly advertised by Hempel in “The Function of General Laws in History” to historiography. The covering-law account states that an explanation of a singular fact is a valid deductive argument where the *explanandum* is deduced from and shown to follow from the initial conditions $C_{1...}C_n$ in accordance with the law statements $L_1 \ldots L_n$ (Hempel 1942: 36; Hempel & Oppenheim 1998[1948]: 207–209). Since the explanatory potential of suggested explanations relies on the existence of laws from which the *explanandum* is deduced with the help of the initial conditions, all explanations that do not explicitly state laws are seen as incomplete explanation sketches (Hempel 1942: 42). For instance, the covering law account denounces the singular causal claim ‘The extinction of the dinosaurs was caused by a strike of a huge meteorite’ as an elliptical explanation (Hempel 1965: 415) since it does not explicitly rely on any laws.

The covering law model immediately seems to raise at least three worries to any special science that deals with token explanations. First of all, it seems that the meteorite strike explains the extinction of the dinosaurs without any reference to laws. Secondly, one can also ask what kind of laws each special science should use. Laws of physics? Are there biological laws? Does historiography have laws of its own or should historians borrow their explanatory laws from social and behavioral sciences? Is the generalization (4.1) ”whenever a nation s subjected to the challenge of the sea, has excellent soil, is assisted by its neighbors, and has an excellent maritime situation, it will rise to the great heights of success” a historiographical law? Is it really needed in

order to explain why the Netherlands rose to prosperity in the 17th century? [13] Thirdly, one can also ask what laws are in the first place.

My purpose is not to dwell on these questions in a detailed manner. They are introduced in order to show that the covering law account seems to have raised more questions concerning the methodology of historiography than it has been able to solve[14]. For instance, were a historian to adopt the covering law model as his view of explanation, it would not give her much of advice on what to do next. The covering law model is committed to the idea that for every causal explanation, be it a token or type level explanation, there underlies a structure of causal nomic connections that gives the causal explanation its explanatory import (Hempel 1965: 360–361). This idea does not, however, give a direct insight on whether the historian should seek to *justify* her singular causal explanations by searching for the covering laws that the singular causal explanation supposedly refers to. As already mentioned, the historian would also be puzzled because she would not be given any advice on where to look for these laws if they are needed: in historiography, social sciences or perhaps, in physics? In the cases where a historian compares competing explanations the advice would be to favor the explanation that has a general law mentioned in it explicitly. The question still remains on how to deal with cases where the competing explanations are equally elliptical?

This methodological unfruitfulness of the covering law model can be seen in Raymond Martin's (1989: Ch. 2) discussion on the covering law model in historiography. Martin starts by presenting the debate on whether historians should try to find covering laws in order to justify their singular causal explanations. He rejects Morton White's and Murray Murphey's claims that covering laws are needed in order to defend historiographical explanations. Despite of this, Martin (1989: 23) still claims that the covering law model is the right account of what counts as an explanation.

On the other hand, Martin (1989: 25) thinks that it is sufficient for a historian to be able to defend her favored explanation against competing explanations in order to rationally endorse it. As an illustration of this idea he discusses the explanatory debate among archeologists and historians on what caused the massive population loss of the

[13] Morton White, who has introduced the generalization (4.1) (1965: 23), thinks that one should be able to defend the generalization in order to defend one's singular causal explanation of the rise of the Netherlands. For more discussion on the topic, see Martin (1989: 19–23). Morris's idea closely resembles the idea that Woodward (2003: 164) calls the epistemological thesis: one needs to know the laws underlying a causal claim if one wants to reliably tell whether the claim is true.

[14] see Kosso (2011: 23–24) and MacDonald & MacDonald (2011: 133–136) for more examples of questions that Hempel's model has raised in historiography.

Classic Period Lowland Maya civilization in the 9th century. But he does not make use of the covering law model anywhere in his analysis in order to make sense of the competing explanations presented in the debate. As already mentioned in the section 4.3, in his discussion of weighted explanations and subjectivity in historiography, Martin adopts the contrastive idea, which has nothing to do with the original covering law idea. In Martin's methodological discussions on explanation in historiography, the covering law model turns out to be a dead letter.

The shortcomings of the covering law model in important methodological questions can been regarded as a consequence of its committal to what Woodward (2003: 159) calls "the hidden structure strategy". Epistemically speaking, a proponent of the hidden structure strategy claims the following: causal claims *explain* in virtue of providing information about the existence of some underlying hidden structures that are the regularities that make up the causal nomic structure of the world (Woodward 2003: 179). The problem is that even if we interpreted the causal structure to be the laws of current physics, most people engaged in causal explanation are not familiar with these laws. In spite of this, people, scientists and historians alike, evaluate causal claims and condemn some explanations bad and good all the time. It is quite implausible to suggest that people do these evaluations based on the causal structure of the world that is epistemically inaccessible to them (Woodward 2003: 179–180). Since in sciences and historiography explanations are constructed and evaluated with the help of experimentation, observations and theoretical knowledge that is not quite often formulated as sets of laws, it is no wonder that the covering law model offers little insights on the questions how things are really explained and different explanations assessed.

At the more general level, the standard philosophical counterexamples introduced against the covering model might also be taken as sufficient in themselves to debunk the covering law model. The various counterexamples prove that the covering law account does not give the sufficient or, worse still, the necessary conditions for an explanation[15]. The reason why I have not extensively presented the counterarguments,

[15] Wesley Salmon's (1971: 34) counterexample of a man, Jones, who takes birth control pills is one example on how the covering law model fails to give sufficient conditions for explanations: it lets the irrelevant fact that Jones is taking birth control pills together with the generalization "all males who take birth control pills regularly fail to get pregnant" to explain why Jones does not become pregnant. Michael Scriven (1959a and 1959b) has extensively argued that the model does not provide necessary conditions for explanations: there are singular causal explanations that do not cite any laws or regularities. For instance, we can explain why Jones developed paresis by referring to his untreated syphilis, although at

but, instead, taken up the methodological unfruitfulness of the covering law model, is that my aim is to reach a more positive conclusion than just debunking the covering law model. The theory of explanation to be applied in philosophy of historiography should have merits of solving some methodological problems instead of giving no real guidance or even creating more methodological controversies. Based on the discussions of this chapter, there are reasons to believe that the contrastive counterfactual theory is this kind of methodologically fruitful theory.

As already stated in the third chapter, the contrastive counterfactual theory is comfortable with singular causal explanations. If a singular causal claim exhibits "systematic pattern of counterfactual dependence" (Woodward 2003: 191), it can be deemed as explanatory. Thus, the contrastive counterfactual model does not require that explanations should be valid deductive arguments and that explanation and prediction should be symmetrical. Nor does it claim that laws are essential to explanations and that explanatory power of singular causal claims is based on references to laws. In other words, the contrastive counterfactual account drops the features that characterize the covering law model on the basis that they do not pick out what is essential to explanatory import: invariance under possible interventions.

However, dropping the central ideas of the covering law model would not be a big merit from a historian's point of view without the methodological insights that contrastive counterfactual account provides. In this chapter we have seen the insights that result from applying the idea of contrastive *explanandum* and contrastive causal claims to the problems of explanatory selection, subjectivity of explanations and negative causes. In addition, the contrastive counterfactual account manages to solve the principled epistemic objection presented by the antagonists of counterfactuals: historiographical evidence can be used to support causal, and therefore, counterfactual claims. The task of the next chapter is to shed more light on the question how this is done in practice. The next two chapters will also show how the account deals with and clarifies the important methodological issues of causal weighting and narrative structure of explanation as well as provides a methodology for counterfactual inferences.

---

the moment medicine is not able to state under which antecedent conditions and together with what generalization, the occurrence of a paresis is highly probable. In addition, only a small percentage of those who have untreated syphilis develop paresis, but the probabilistic version of the covering law model – the IS model – says that the laws and the antecedent conditions should make the *explanandum* highly probable.

5 Epistemology and methodology of counterfactual inferences

One ought to be able to associate with any successful explanation a hypothetical or counterfactual experiment that shows us that and how the manipulation of the factors mentioned in the explanation would be a way of manipulating or altering the phenomenon explained.

– James Woodward (2003: 11)

According to the theory of causal explanation presented in the earlier chapters, causal claims imply counterfactuals. This is not merely a semantic point, but also an epistemological one. If one has causal knowledge, one should also be able to engage in counterfactual reasoning and answer *what-if-things-had-been-different questions* (*what if-questions* for short). This means that the more *what if*-questions one can answer, the more causal knowledge one has. *What if*-questions are not answered without evidence however. Only with proper kind of evidence can one make reliable causal and counterfactual inferences. Thus, it is crucial to have and idea of what kind of evidence historians can use when answering *what if*-questions. How does the evidence that is available to historians limit the certainty of their causal and counterfactual claims?

This chapter begins with the questions of evidence and counterfactual reasoning (section 5.1). The second part of the chapter (section 5.2) explores to what extent the contrastive counterfactual theory and the idea of interventionist counterfactuals can help to solve some of the most pressing problems of counterfactual inference that historians have presented. The point is to shed some light on the issue of what kind of counterfactuals to use and what to dismiss. This scrutiny also identifies the reasons why counterfactual inferences sometimes seem to be insurmountable for historians. In addition, the scrutiny reveals the difference between counterfactual inferences on *what could have happened* and inferences on *what would have happened.* The latter inferences are central for causal reasoning, whereas the first ones are not.

5.1 Epistemology: Counterfactual reasoning and the use of evidence

When discussing counterfactual and causal reasoning, many current protagonists of counterfactuals talk of "counterfactual thought experiments" (e.g Tetlock & Belkin 1996; de May and Weber 2003; Lebow 2010: ch.2). When conducting an experiment one applies the method of varying one variable and leaving the values of other variables

intact in order to see what kind of effect this change has on the outcome of interest. When conducting a counterfactual thought experiment the varying is primarily done, as the term 'thought experiment" suggests in the mind of the historian[1].

The terminology of "counterfactual thought experiments" is slightly problematic, however, since the talk of "experiments" does not reveal that there is a crucial difference between the evidential roles of real experiments and thought experiments. Real experiments in the sciences are conducted by actually controlling for other variables in order to find out the effects of certain variables. These experiments generate evidence that can be used to support causal hypotheses and to discriminate between competing hypotheses. Instead, *counterfactual thought experiments do not generate any new empirical evidence*. Counterfactual thought experiments are inferences in which one makes use of the readily existing evidence in order to support one's claims of what would have happened if this and this variable had been different in a certain way. In other words, counterfactual thought experiments cannot be used as independent evidence to support causal hypotheses. They are means to explicate and to scrutinize the implications of causal hypotheses and they provide a means to check how much support a certain causal claim has in the light of the evidence. This evidential impotency is a crucial difference between actual experiments and counterfactual thought experiments, and therefore, in order to avoid confusion I talk of *counterfactual inferences* instead of counterfactual thought experiments.

So, what kind of evidence historians make use of when assessing causal claims and the related counterfactual inferences? Historians are not, in general, in the position to conduct controlled experiments, which in the light of the theory of causal explanation defended here do seem to provide the optimal, although by no means perfect, way of answering *what if*-questions. In what follows, I will sketch one method that is useful for causal reasoning and answering *what-if*-questions in historiography. Due to lack of space I cannot focus on statistical inferences and statistical methods, which in some fields of history like in economic or population history are a valuable tool for making causal and counterfactual inferences[2]. Nor can I discuss the exotic methods like the role of computer simulations in causally explaining the past[3]. Instead I will discuss the

[1] This does not mean that external aids like pen and paper or even computer simulations could not be used to help out in these inferences.
[2] On the use of statististical methods in historiography see e.g. Jarausch & Hardy (1991) or Feinstein & Thomas (2002).
[3] On the use of computer simulations in counterfactual inferences, see Cederman 1996.

method of process tracing in the context of qualitative case comparison studies and single-case studies[4].

Simply put, process tracing is about finding the mechanism(s) that produced the outcome of interest (Steel 2008: ch.9). Mechanisms are usually considered to be repeatable or frequently occurring combinations of entities and their organizations that interact according to some invariant causal generalizations to produce the outcome of interest (see e.g. Hedström & Ylikoski 2010 for a general overview on the definitions of mechanisms presented in the literature). In some fields of historiography the idea of mechanisms as *repeatable* is easier to grasp and to accept than in some others. For instance, in economic history it is common to suppose that the same kinds of market mechanisms have occurred in various contexts at different times to produce similar kinds of outcomes. However, in other fields of history the mechanisms that historians are interested in are considered to be local, highly context-dependent or as Stuart Glennan (2010) puts it: ephemeral. Therefore, repeatability is not a property that mechanisms studied in historiography necessarily have: a mechanism that a historian is interested in can also be a one-time combination of entities and their organizations that interacted according to some invariant causal generalizations[5] to produce the outcome of interest (ibid). Let's discuss two examples to highlight the ways in which process tracing works in comparative case studies and single case studies to provide evidence for causal and counterfactual claims.

Katalin Miklóssy (2007) presents a comparative case study on Hungary's and Romania's changing roles in the international politics during the Cold War. The comparison begins from the year 1956 and ends in the late 1970's, when the consequences of the OSCE-meetings in Helsinki (1975) and Geneva (1973-1975) had become apparent. Miklóssy asks why Romania managed to lose its good reputation in the minds of the Western leaders during 1970's, whereas Hungary managed to leave behind the isolationist consequences of the uprising of 1956 and to become the West's favorite socialist country during the OSCE-process? In the 1960's Romania was able to

[4] The methods mentioned are not mutually exclusive; they can all be used to together in order to arrive at conclusions on causal relations.

[5] The term 'generalization' does not only refer to patterns that repeat in various different context. Invariant causal generalizations can also be context-specific: they describe how things work or are done in general in a certain context. Therefore, also social rules of conduct that tell how people or certain inviduals behave in some cultural or social contexts are also interpreted as invariant causal generalizations. They can be used to support causal claims: the brothers-in-law in the Trobriand tribe give presents to the husbands of their sisters, because brothers-in-law [instead of grandfathers or sisters-in-law] are expected to give presents to the husbands of their sisters.

break the leash of Soviet control by conducting independent economic and trace policies and it became the first “satellite country” to form official trade relations to the west. During the 1970’s, however, both the Soviet Union and the western countries turned their backs on Romania. Hungary, on the other hand, abandoned its radical economic reforms in 1968 in the fear of another military intervention by the Soviet Union. Despite of this drawback in the late 1960’s, Hungary managed to quietly restructure its economy and and to form 55 new bilateral trade agreements with western countries during the years 1976–1977.

Miklóssy’s method of finding the factors that account for this difference is to compare Hungary’s and Romania’s foreign policies, and to point out the differences in these policies that in the changing international environment lead to different kinds of outcomes. Her conclusion is that Hungary was able to gain a good reputation in the west by stressing the importance of human and minority rights issues in international arenas in the 1970’s. The 1970’s was the period when the western countries also raised these issues onto their political agenda. This new emphasis on human rights reflected negatively to Romania’s reputation: Ceausescu’s Romania came to be known as the brutal totalitarian regime suppressing the rights of its citizens in a Stalinist manner. The loss of reputation was the cause why the western countries decided not to renew their trade agreements with the country.

Miklóssy’s explanation for why Hungary managed to keep up good relations both to Soviet Union and to the West, was its tacit way of operating in international arenas. Officially, Hungary followed the lead of the Soviet Union, for instance, by backing up the occupation of Czechoslovakia in 1968. This manner of conducting gave Hungary enough space to quietly fulfill its own policy goals. Ceausescu’s loudness in stressing the independence of Romania annoyed the Soviet leaders to the extent that when the western countries turned their backs on Romania, the Soviet Union was not there to help out.

Miklóssy’s conclusions give support to several counterfactual claims, although she herself does not formulate her results in counterfactual terms. For instance, one can conclude as follows: if Romania had managed to play by the rules set by the Western countries and the Soviet Union in the same way that Hungary did, it is likely that the country would not have ended up as an outcast in international politics. The factual support for this counterfactual comes from the contrast between Hungary’s and Romania’s conduct in international arenas and the different outcomes that the differing

ways of conducting led to. Hungary played by the rules set by the more powerful ones and was successful in quietly transforming some of the rules to suit itself, whereas Romania wanted to play by its own rules and was left in the position where it was in no position to change to rules of those with the most power.

What makes Miklóssy's comparative case study an example of comparative process tracing? First of all, the study is motivated by the observation that Romania and Hungary were both small socialist countries that despite the somewhat similar starting points ended up occupying different roles in the field of international relations. The idea is to explain the difference of outcome by comparing the mechanisms M_1 and M_2 that lead from the similarity of the starting point A to the difference in the outcome B [Hungary's popularity in international arenas; Romania's unpopularity]. In other words, what is of interest in process tracing, is the causal relation between A and B, and the causal and counterfactual inferences on the relation between A and B are supported by the evidence resulting from exploring mechanism(s) M_n that link A and B together. The assumption is that the causal relations within the mechanisms are easier to find or to infer than it is to directly infer the holding or the nature of the causal relation between A and B (Steel 2008: 187). Therefore, the mechanisms linking A and B are the means to infer what is going on between A and B and thus the knowledge on the mechanisms can be used to support causal claims on the relation between A and B and to explain B.

What is needed for the process tracing to work is[6] identifying i) *the entities of the mechanisms* = various countries in the field of international relations, most notably Romania and Hungary, the Soviet Union and ii) *their organization* = the division of power between small countries and superpowers iii) *the invariant causal generalizations that guided the interactions among the organized entities* = the "rules" according to which international relations operated in the 1960's and 1970's that defined what small countries and superpowers were able to do[7]. Comparing the mechanisms M_1 and M_2 where the two countries reacted in different ways to the "rules" of the game supports causal and counterfactual inferences that explain the contrast in B [Hungary's popularity in international arenas; Romania's unpopularity].

Careful process tracing on *a single case* can also give a historian enough knowledge to do counterfactual inferences and to compare the actual case with the

[6] See Steel 2008: ch. 9 for a detailed analysis on how process tracing works.
[7] These rules did not stay the same during the years 1955–1980; were in constant change due to the interactions between entities and it was precisely these changes that affected Hungary's and Romania's positions in international relations.

inferred counterfactual case (and in this way to compare cases). I use Steel's (2008: 189) discussion on Malinowski's (1935) study on the wealth and influence of Trobriand chiefs to illustrate this. Malinowski's causal hypothesis was that the possession of many wives was the cause of wealth and influence among Trobriand chiefs in the early 20th century. Based on this causal hypothesis one can also make the following counterfactual claim: if a certain chief had had more wives, he would have been richer. According to Steel (2008: 189), Malinowski supports his hypothesis between the number of wives and wealth by tracing the social mechanisms that link these two together.

The process tracing on the causes of the Wealth of the Trobriand chiefs reveals i) actors (the entities of the mechanism) that are ii) organized in accordance to the social roles attributed to them (e.g. chiefs, wives and brothers-in-law) and iii) invariant causal generalizations that govern the interactions between the actors that are organized based on their roles: the social rule is that brothers-in-law are supposed to give presents to the husbands of their sisters – the gift is larger when the sister is married to a chief and the chiefs usually have many wives. The knowledge provided by i), ii) and iii) supports Malinowski's causal claim. This example is taken from cultural anthropology, but historians' single-case studies also often rely on this kind of methodology of process tracing: the aim is to explain why a certain thing happened in this way instead of some other way by tracing the causal mechanisms that worked in a certain historical context and accounted for the contrast of interest.

As the reader might have noticed, this section has contained a lot discussion of evidence without clearly specifying what is this evidence that historians lean upon in justifying their causal claims. Process tracing is one way to use evidence to support causal and counterfactual claims, but what kind of evidence is needed in process tracing? What kind of evidence is used in historiographical causal reasoning in general? First of all, historians' *evidence* consists of *traces of the past*. The traces can be anything that has the potential to tell something of the time of interest[8]. For the purposes of process tracing, a historian needs traces of the past that help her to infer i) what the central actors and entities were ii) how they were organized iii) what kind invariant causal generalizations there were to guide the interaction between the actors and other entities. Ideally, knowledge of the mechanisms that were present at the time of interest

[8] "Source critisism" is the method that historians use to check whether the source dates actually from the time of interest and gives reliable information of the issue of interest.

helps historian to formulate and also to support various causal and counterfactual claims.

Knowledge of the mechanisms that is needed in process tracing to support causal claims does not merely stem from the traces of the past. A historian can also apply her *general theoretical knowledge* of how societies, markets and humans behave into the case one is studying in order to make causal inferences. This kind of general knowledge that is applied into the cases that a historian is working on form *the other major source of evidence* for a historian. General theoretical knowledge can be used to supplement the knowledge inferred from the traces of the past or to interpret the traces of the past in order to get a fuller picture of the causal relations and mechanisms of the past. When a historian applies her general theoretical knowledge, she supposes that the same mechanisms can work in variety of different contexts to produce similar outcomes. For instance, the applications of rational choice theories to various historical cases are an example of this kind of applying (see e.g. Bates et al. 1998).

Historians often have also various implicit theoretical ideas on invariant causal generalizations and on how things work. They often also implicitly rely on these ideas in reconstructing and inferring causal relations of the past[9]. Historians' theoretical ideas often stem from the views presented in the earlier history writing. They can also be based on the theories of society and human behavior from various social and behavioral sciences, as well as, the statistical and qualitative studies from these sciences. In addition, historians quite often make use of their everyday beliefs about how do people behave. When these ideas are not spelled out explicitly, the they are applied implicitly: one makes causal claims on the past causal relations without specifying all the evidence one uses to support these claims.

However, being serious about causal and counterfactual reasoning means that one should be as explicit as possible about the evidence that one has in support of one's claims, since the "strength" of the evidence determines the reliability of the counterfactual inferences. If one wants to do serious work on the past causal relations one needs to spell out not only the evidence that stems from the time of interest, but also where one's general theoretical ideas of the working of societies and behavior of humans come from – at least in cases where these theoretical ideas shape the causal

[9] The prominent Finnish historian on international relations, Juhana Aunesluoma, (2010: 64–65) describes the implicit use of theories in historiography as the method where "the theories are baked inside the narratives".

claims that one makes. One should not hide behind "historian's imagination" that helps one to make coherent interpretations of the past causal relations, but rather try to document the evidence that one has in support (and against) certain causal claims. By explicitly engaging in counterfactual reasoning and by representing the evidence in favor of counterfactual inferences is an efficient way to try to see how much evidence one has in support of one's causal claims. Being explicit also makes it easier also for other historians to intersubjectively to assess the quality of one's work.

The fact that the justification of counterfactual inferences depends on the evidence that there is to support the inference, reminds us of the epistemological doubts of the antagonists of counterfactuals that I presented in the second chapter. *The modest epistemic objection* says that there might be cases where the lack of evidence makes it impossible to assess the truth of the proposed counterfactuals. The classic example of Cleopatra's nose serves as a nice example of this point. Is the counterfactual "Cleopatra's nose was the cause of the World War I" (see: Fearon 1996) to be deemed as true or untrue[10]? My answer is that based on the evidence that we have at the moment, we do not know. We simply do not have enough evidence on the causal mechanisms from the time of Cleopatra to the beginning of the First World War. Some of us might have some metaphysical ideas of the nature of history as a process where everything affects everything, which would mean that Cleopatra's nose would be the cause of everything that has happened after the time of Cleopatra. On the other hand, some of us might have intuitions that perhaps everything does not affect everything. Cleopatra's nose might have affected the behavior Mark Antony and Julius Caesar and the wars they started, but perhaps these wars did not make a big difference in the long run. In any case, the truth of the counterfactual "Cleopatra's nose was the cause of the World War I" is an empirical question. This is why we should not base our assessment of this counterfactual merely on our metaphysical intuitions, although we would lack sufficient empirical resources to reach any firm conclusions on the matter.

An antagonist of counterfactuals is surely right in pointing out that the lack of evidence might prevent us from deciding on the truth-value of some counterfactuals. For some counterfactual inferences, on the other hand, the situation might not be that despairing. Absolute certainty cannot be expected from counterfactual inferences, since

[10] Blaise Pascal introduced the well-known example of Cleopatra's nose in his *Pensées* with the counterfactual: "Cleopatra's nose, had it been shorter, the whole face of the world would have been changed" (*Pensées* II, 162).

absolute certainty is out of the reach of any empirical science. So there is always some degree of uncertainty in every counterfactual inference. By careful use of evidence, one can try to assess just how reliable one's counterfactual inferences are. This is the only way to avoid the pitfalls of presenting one's counterfactuals more strongly supported than they are and to separate fanciful "iffing" from serious empirical work.

5.2 Interventionist methodology for counterfactual inferences

One source of problems for a historian working with causal explanation and counterfactual claims is the fact that evidence can be scarce, shaky and non-decisive. Other bundle of problems descends from the difficulties of deciding on how to go about with counterfactual inferences. What kind of things should a historian focus on, and what kind of things to neglect when one tries to determine the truth of her counterfactuals with the help of evidence? This section deals with this question by looking more closely to the properties of interventionist counterfactuals.

5.2.1 Features of interventions

Raymond Aron (1961: 160–161) provides a general illustration of the general idea of how one should proceed in counterfactual inferences when one wants to determine the causes of events:

STEP 1. Analysis of the phenomenon-effect.

STEP 2. Discrimination of the antecedents and isolation of one antecedent, the efficiency of which it is desired to measure.

STEP 3. Construction of unreal developments.

STEP 4. Comparison of mental images and the actual events.

In the light of the contrastive counterfactual theory Aron's idea could be interpreted as follows:

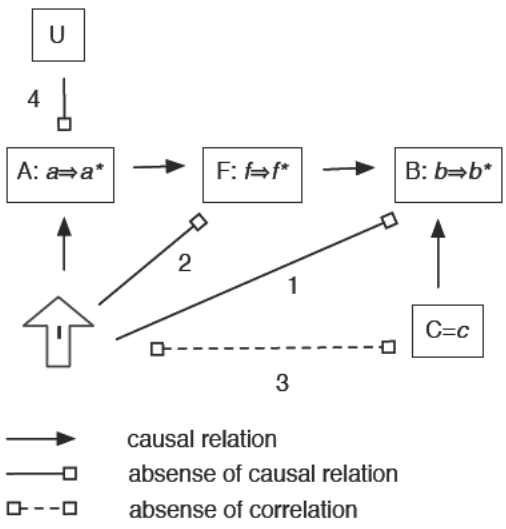
STEP 1.* Analysis of the phenomenon-effect = define the contrastive *explanandum* *b[b*]*.

STEP 2.* Discrimination of the antecedents and isolation of one antecedent, the efficiency of which it is desired to measure = If the fact *a[a*]* is hypothesized to be the explanatory cause of *b[b*]*, imagine the intervention that changes *a* to *a** and leave the other causal routes that lead to the variable B intact.

STEP 3.* Construction of unreal developments = specify on what evidential basis one can expect that *a** leads to or alternatively does not lead to *b**.

STEP 4.* Comparison of mental images and the actual events = the comparison is done between what is known based on the evidence to have happened and what the evidence justifies us to believe what would have happened, if there had been an intervention on *a*.

This four-step counterfactual inference (or counterfactual thought experiment, if you like) is meant to "test" whether there is an invariant relation between *a[a*]* and *b[b*]*. This is done by hypothesizing an intervention from *a to a** that only affects the values of those variables that are on the causal route from *a* to *b*, and keep the values of other variables V that are not on this route as they actually were $v_1 \ldots v_n$ (Woodward 2003: 77). In chapter 3, we learned that an intervention on some *explanans* variable A with respect to the *explanandum* variable B is a causal process, "a switch", that changes the value of A in an appropriately exogenous way, so that if a change in the value of B occurs, it occurs only in virtue of the change in the value of A (Woodward 2003: 94). However for the task at hand one needs to be even more precise on the nature of interventions. Picture 5.1. shows the conditions *that an ideal intervention* from *a* to *a** should fulfill (The graph is adapted from Craver 2007: 97 and Ylikoski 2011: 166. See also Woodward 2003: 95–98).



1. I does not change B directly.

2. I does not change the value of any causal intermediate F between A and B except by changing *a* to *a**.

3. I is not correlated with some other variable C that is a cause of B.

4. I acts as a switch that controls the change from *a* to *a** irrespective of A's other causes U.

Picture 5.1. Ideal intervention and its conditions

It is important to note that these conditions describe *ideal* interventions and *ideal* interventionist counterfactuals. It might be that no intervention carried out in scientific experiments or no counterfactual inference in historiography is able to fulfill all these conditions. This does not mean that interventions that fall somewhat short of fulfilling these conditions would not give any causal knowledge. The conditions of ideal intervention can be seen as an regulative ideal (Woodward 2003: 114): they tell what we should aim at in our investigation, since they tell what must be true in the relationship between A and B if A causes B.

In what follows I will go through the conditions one by one in order to see to what extent historians' counterfactuals fulfill these conditions. I will analyze why certain counterfactuals fail to fulfill these conditions and explore the consequences of these failures. Even if a counterfactual fails to fulfill these conditions, it does not mean that the counterfactual should be dismissed out of hand. The analysis in section 5.2.2 reveals that, in fact, there is no algorithm that would help us easily to discern counterfactuals with "dubious" antecedents from counterfactuals with more respectable antecedents. Sections 5.2.3 and 5.2.4 argue that sometimes historians' object of study – the causal relations in the past – makes it practically impossible for a historian to conduct counterfactual inferences that fulfill the conditions 2 and 3. It is hard to make inferences that fulfill the conditions 2 and 3 if the causal system under study resists to work in accordance with these conditions: complex systems quite often require complex inferences, and non-modular systems often lead to non-modular inferences.

Section 5.2.5 deals with the condition 4, and argues why it is always sensible to strive to make counterfactual inferences that fulfill this condition if the aim is to make causal inferences on the relation between A and B. Historians quite often make counterfactual inferences that do not fulfill this condition, and it is argued that that these inferences do not contribute to causal reasoning in any straightforward way. In the end, the analyses of sections 5.2.2–5.2.5 sketch a picture of what historians causal reasoning is like, and compare it to how it should be like. With what kind of inferences is it reasonable for a historian to explore causal and explanatory relations?

5.2.2 Intervention does not change the *explanandum* variable directly: the problem of selecting sensible causes

The need for a condition 1 (*the intervention I does not change the variable B directly*) is quite easy to understand: if we want to test whether A is the cause of B, it makes no sense that the intervention would change B directly. A good example of an intervention that does not fulfill the condition 1 is a randomized drug experiment where the subjects in the treatment and control groups are let to know whether they are treated with the drug or with a placebo and where this knowledge of the intervention affects their probability of recovery from their condition. Are there interventions and counterfactual inferences that do not fulfill this condition in historiography? This question turns out to be a tricky one, since it is not obviously clear how the condition 1 should be interpreted.

One group of counterfactuals that do not seem to fulfill this condition are those where changing the antecedent from *a* to *a** would "annihilate" the contrast *b*[*b**] altogether. Two following two counterfactuals seem to work this way at first sight:

(5.1) If Romania had been a superpower like the Soviet Union and the United States [in contrast to being the small country it was], Romania would have been able to resign from the Warsaw pact in the 1970's [instead it stayed a member of the alliance although it tried not to take part in its operation].

(5.2) If Napoleon had had Maxim machine guns in the battle of Waterloo [which he did not have], he would have won [instead of losing][11].

How to treat these counterfactuals? According to one line of thinking the counterfactual inference would go like this: let's imagine that we do an intervention where we change the status of Romania from a small East block country to a superpower. This intervention means changing the history of the country significantly in such a way that the country would not have become a member of pact dictated by the Soviet lead in the first place. Thus, the contrast implied in the consequent of this counterfactual would be have been left unrealized and the whole question of Romania's status in the Warsaw pact become pointless if the antecedent had been realized. The antecedent and the consequent do not seem to fit together: if the antecedent realizes, there are reasons to believe that the consequent cannot realize. The same applies in the case of the

[11] This example is adapted from Fearon (1996: 40). Maxim machine gun was invented in 1885 by Sir Hiram Maxim and was used extensively in the First World War.

counterfactual (5.2). If the situation had actually been so that Maxim machine guns existed in the beginning of the 19th century, the situation in international politics would most likely been altered in a way that the Battle of Waterloo would not have been fought.

Many protagonists of counterfactuals who have dealt with similar kinds of tricky counterfactuals, have rejected counterfactuals resembling (5.1) and (5.2) by arguing along the lines presented above (e.g Tetlock & Belkin 1996, Fearon 1996: 57). This type of reasoning seems very plausible, but it is not the only possible way to go about interpreting counterfactuals (5.1) and (5.2). The condition 4 set for ideal interventions states that interventionist counterfactuals do "miracles" in the sense that they break all the links between the causes of the antecedent (the variables U in the picture 5.1) and the antecedent A. The point is to *isolate the relation between A and B so that this relation, and this relation alone,* can by studied. We want to know what kind of effect A, and A alone, has on B is and for this purpose we want to manipulate A, not the causal history of A, with the help of an exogenous intervention.

Therefore, according to the condition 4, when we are thinking of possible interventions we should not concern ourselves greatly with the questions on how the interventions could actually have realized or could they have been realized. I will argue further why it is better to interpret counterfactuals along the lines of the condition 4 in the section 5.2.5. But for now, however, in order to pursue further with the topics of this section, let's explore how the counterfactuals (5.1) and (5.2) are be interpreted if we follow the condition 4.

In order to imagine an intervention in the international status of Romania, we do not need to rewrite and to imagine to whole history of the 20th century and to create an alternative history where Romania becomes a superpower by the 1970's and to speculate whether Romania would have become a member of the Warsaw pact in this process or not. What we, according to the condition 4, need to do is to postulate that Romania would have suddenly become a superpower with the central properties of superpowers[12] in one night, let's say, in 1975, and ask whether this sudden change of status would have helped Romania to retreat from the pact. Based on what we know of the dynamics of international relations in the 1970's, it seems to plausible to suppose that a superpower could have resigned from pacts like the Warsaw pact or NATO. So, if

[12] E.g. economic, military, technological, cultural leadership and due to this factors the power to work in the field of international relations relatively independently.

we interpret the counterfactual (5.1) in this manner, it is not pointless, senseless or absurd – it is, given the things we know, most likely true. This result applies also to the counterfactual (5.2). If Maxim machine guns had suddenly fallen from the sky and Napoleon's troops had suddenly known how to use them, they would have won the battle of Waterloo.

If we belief that counterfactuals (5.1) and (5.2) reflect invariant dependency relations, it means that the fact that "Romania was not a superpower in the 1970's" should be regarded as a cause for the fact that "Romania did not break free from the Warsaw pact". This sentence might be considered slightly odd at first, but when we realize that "Romania was not a superpower in the 1970's" means in this context the same as the sentence "Romania was a small country", the counterfactual inference becomes more comprehensible: the fact that Romania was a small country instead of being a superpower explains why Romania was not able to break free from the Warsaw pact.

The fact that "Napoleon did not have Maxim machine guns" is also to be considered a cause for his defeat at Waterloo. There is a right kind of invariance relation between the intervention on the variable of Napoleon's field weaponry [Maxim machine guns; no Maxim machine guns] and the outcome of the battle [victory: defeat]. Despite there is a right kind of counterfactual dependency relation between the variables, we do not find it that sensible to say that the lack of machine caused the defeat. Why is this? Why does the fact that Romania was a small country strike as a more sensible cause than the fact that Napoleon did not have Maxim machine guns?

There are several other examples that resemble the Napoleon case. James Woodward (2006: 24–25) discusses the following example: the counterfactual (5.3) "if I was struck by a huge meteorite, I would not sit here writing this thesis" is apparently true. Therefore, the fact that I am not struck by a huge meteorite explains why I am sitting here writing this thesis. To further expand Woodward's example the fact that the Earth has not been struck by a huge meteorite seems to have been a cause to everything that has happened within the past 65 million years or so.

The counterfactuals (5.2) and (5.3) point out an interesting problem in the contrastive counterfactual theory of causal explanation. The above-mentioned counterfactuals present invariant dependency relations, and therefore according to the contrastive counterfactual theory we should consider them as causes. On the other hand, we are quite unwilling to consider the lack of certain type of machine guns and not

being struck be a meteorite as causes. How come? The reason is not that these causes are negative causes that cite omissions. As we saw in the previous chapter omissions can reasonably be considered as causes in some cases.

The problem has not been left unnoticed among the supporters of the contrastive counterfactual theory. Woodward (2003: 90) answers with the following remark: "Causal judgments reflect both objective patterns of counterfactual dependence and which possibilities are taken seriously." In other words, we consider some variables and their values as sensible or plausible causes, whereas other variables and their values we reject as implausible and these considerations enter into our judgments on which causes we are willing to take seriously and to present as causes[13].

So, what informs our considerations on which causes are to be taken seriously and which not? Why are not Maxim machine guns and meteorites sensible causes although we belief that some interventions on these variables have the right sort of counterfactual dependence the effect variables? Why is the fact that Romania was a small country a sensible cause variable? There seems not to be clear criteria for deciding which causes to count as sensible and which not. There can be some heuristic guidelines, however. The variable of international influence that can take the rough values [superpower; small country] seems to be a sensible cause variable since there is evidence that

i) This cause variable existed in 1970's and it took different values = countries were labeled as superpowers, small countries, middle-sized countries etc. based on their ability to affect the direction of international politics
ii) The field of international relations was organized so that the cause variable worked (and still works) as a difference-maker in international politics: there were mechanisms (the way the relations between countries were organized and the rules that governed these relations) that ensured that possible interventions on or changes in the cause variable would have made a difference to certain effect variables.

[13] This is not a problem that raises merely in the context of historiographical counterfactuals. This is a general problem of the interventionist theory (Woodward 2003: 89), and something like the notion of "serious possibility" is needed to explain why not all counterfactuals that correctly reflect the invariant dependency relations, are considered as explanatory.

Based on this example, we can generalize the following criteria that seem to affect our considerations of sensibility:

(C1) *the existence criterion*: the variable needs exist in the case under study or if the variable exists, the suggested value of the variable needs to have existed.

(C2) *the mechanistic criterion*: there needs to be a mechanism or mechanisms that ensure that (in principle at least) the variable works as a difference maker. In other words we need to have an idea how the variable would work as a difference maker in a certain context[14].

So why are Maxim machine guns not a plausible cause for the Napoleon's defeat? The answer is simple: Maxim machine guns did not exist at the time of Waterloo: the variable concerning Napoleon's field weaponry cannot be given the value of maxim machine guns. The existence criterion is not fulfilled. The meteorite example at first sight seems to fulfill both the criteria (C1) and (C2): meteorites surely exist and we have an idea how the falling meteorites can cause the damage they do: if I was struck by a meteorite I would most likely die, since the mere speed of the collision would break vital parts of my body. The problem with the contrastive *explanans* [I am struck with a meteorite; I am not struck with a meteorite] is that it doesn't seem to "hit" the contrastive *explanandum* [I am writing my thesis; I am not writing my thesis] in a right way. First of all, the contrastive *explanans* explains several other contrasts as well: [why my lungs are working; why they are not working] or [why I am alive; why I am not dead]. However, the *explanans* seems to fair rather poorly in explaining these contrasts as well. What is lacking is the context that links the *explanans* and the *explanandum* together in a right way. The criterion (C2) is not fulfilled after all.

Let us suppose that I have read in the newspaper this morning that there will be huge meteorite showers in Helsinki this afternoon. Before reading the news on weather, I was determined to have a day off and not to write my thesis. Reading the news, however, made me change my mind. I thought that perhaps it is better to write and finish my thesis this morning before the meteorite storm hits me or the room I use to write. In this kind of scenario the explanans [I am struck with a meteorite; I am not

[14] This criterion resembles Petri Ylikoski's (2001: 41) mechanistic requirement: "*a* explains *f* [*c*] if and only if (1) *a* belongs to a causal process leading to *f*, and (2) there is *a* causal mechanism *m* that ensures that *f* occurs instead of *c* because of *a*". The mechanistic condition also ensures that there is a chance to do process tracing to see whether the suggested cause and the explanans are connected or not and if they are in what ways.

struck with a meteorite] seems to be one plausible explanation the contrast [I am writing my thesis; I am not writing my thesis]. This means that our background knowledge of the context defines our considerations what to take as plausible causes and what not. The knowledge on the context defines our beliefs whether there was a mechanism that ensured that the suggested variable or its values worked as a difference maker in a right way.

This result should not be of great concern to historians, however. It is true that we select from the vast range invariant counterfactual dependency relations those relations that we belief on the basis of our evidence to fulfill also the mechanistic criterion. This means that historians should keep on doing what the have been doing: they should try to analyze their evidence, the sources of the past and theories on how things work, in order to create hypotheses about the causes that in the context under study could have brought about certain outcomes of interest. Careful study of the historical context of interest helps to postulate causes that fulfill the criterion (5.2) without any further ado. If the evidence does not imply anything of the possibility of Maxim machine guns or meteorite strikes in the context of the Battle of the Waterloo, there is no point to hypothesize about these variables as causes for the defeat at Waterloo.

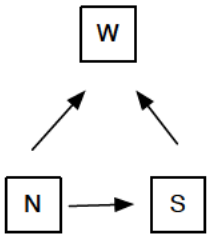
Philosophers who might be interested in finding out clear-cut criteria to separate sensible causes from the non-sensible ones, might not be too pleased with the analysis I have given[15]. In any case, the above-mentioned criteria are at the moment the deepest level of analysis I can go when trying to make sense why certain causes are not generally considered as causes, and I believe that in the context of historiography they suffice. The main point of this section should be clear, however: our causal judgments on sensible causes do not entirely follow the objective invariant dependency relations reflected by counterfactuals interpreted along the lines of conditions 1–4. Secondly, the framework of contrastive counterfactual theory does not collapse entirely in the face of this problem. Instead it helps us to get to the point where we can see in a brighter light where the problem lies: it is not so, that counterfactuals with problematic antecedents should be considered meaningless, as it often happens when we try to rewrite the historical context so that the realization of the antecedent becomes plausible. Instead, many counterfactuals with problematic antecedents can be said to have truth-values if

[15] Also some historians and social scientists express the need to have this kind of criteria see. e.g Fearon 1996: 40–41 ; Tetlock & Belkin 1996: 23–24

we interpret them along the lines of the condition 4 – the problem is just that not every true (or untrue) counterfactual cites a cause variable that is of great interest to a historian.

5.2.3 Intervention should not affect "the middlemen" independently of changing A: the problems of complexity and non-modularity

It is quite easy to understand why ideal interventions fulfill the condition 2 (*I does not change the value of any causal intermediate F between A and B except by changing a to a**). We want to know how a certain intervention on A affects B, not what the total effect of the intervention that affects directly both A and "the middleman variable" F is. For instance, let's go back to the case study by Malinowski on the Trobriand society, introduced by Steel (2008: 189). Malinowski's hypothesis is that the number of wives (N) affects the wealth (W) of the chiefs. There is also, a third variable central to the accumulation of wealth. That is social status of the chiefs (S). Based on Malinowski's case study it is not possible to infer what the relationship of the three variables to each other actually was (Steel: 193 –194), but let us suppose that (i) the number of wives (N) affects wealth (W), ii) the number of wives (N) affects social status (S), but does not totally determine the value of (S): there are other factors that also affect (S) (iii) the social status (S) affects wealth (W). The causal system looks thus as follows:

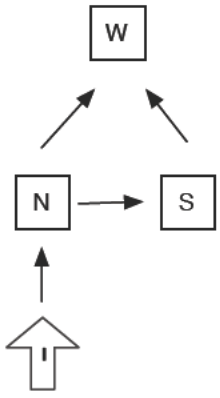


Picture 5.2. The number of wives (N), social status (S) and wealth (W)

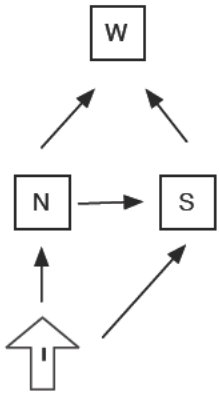
This causal system was intervened on by the British rule during the early 20th century. The British banned the polygamy based on moral reasons (Steel 2008: 194). Let us imagine an anthropologist or historian interested studying Malinowki's hypothesis. For her the British intervention might seem like a great opportunity to explore the truth of the hypothesis. If the historian has evidence on the wealth of the chiefs from the time before the intervention and from the time after the intervention, she

is able to infer what kind of effect the number of wives had on the wealth of the chiefs. If there is a decrease in the wealth, the number of wives had a positive role in the accumulation of wealth. If there is an increase, the number of wives had a negative effect. The success of these inferences depends on the assumption that the British intervened solely on the number of wives (N).

In reality however, the British rule intervened on the social status of the chiefs via other means than by banning polygamy. The aim was to undermine the position of the chiefs by implementing various other laws and orders, and thus give more power to the British colonial authorities (Steel 2008:194). In other words, in reality, the British intervened on (S) also directly, not only through the intervention on (N). This is bad news for our imaginary historian or anthropologists. Since (S) was also affected directly, it is not possible to distinguish the overall effect of the number of wives (N) on the wealth of the chiefs (W). The historian is left to conclude that together the number of wives and social status had (presumably) a positive effect on the wealth of the chiefs – it might even be that (W) was causally only dependent on (S), whereas (N) did not have any effect on (W) neither directly or indirectly through (S) unlike the picture 5.2 suggests. When (N) and (S) are both intervened simultaneously, one cannot even infer whether the causal structure of the picture 5.2 is the right causal structure or not. This example illustrates why it is important that interventions do not affect the "middlemen variables" that are between the *explanans* variable of interest and the *explanandum* variable. If the interventions affect the middlemen variables directly, one cannot reliably infer what the relation between the *explanans* variable and the *explanandum* variable was or was there a relation to begin with.



Picture 5.3. British intervention on N



Picture 5.4. Actual British intervention on S and N

The methodology of intervening on just one variable and keeping other things equal was not possible for our poor imaginary anthropologist or historian because intervention by the British colonial rule proved to be a non-ideal one. Is the methodology possible in historiography in other cases, however? Can there be counterfactual inferences that neatly fulfill the condition 2? These questions cannot be given a definite answer, since it depends on the nature of causal structure under investigation what kind of interventions are possible. So the right kind of question is what kinds of causal systems or networks of causal connections historians are interested in. It seems that in many cases historians are interested in systems where it is hard to intervene on some causal connections without affecting the other causal relations. Richard Ned Lebow (2010: 56–57) puts the point as follows: "History is like a spring mattress. If one spring is cut, or simply subjected to extra pressure, the others will to varying degrees shift their location and tension".

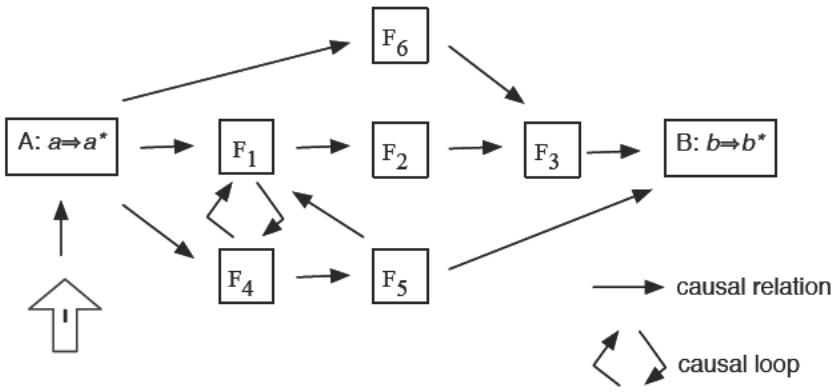
How should we interpret Lebow's comment? At this point it is important to distinguish between two properties that systems can have that both reflect on our ability to do counterfactual inferences on these systems: *complexity and modularity.* Failures in fulfilling the condition 2 are failures because they jeopardize the *modularity* of the intervened causal system: if a system is modular (or our representation of the system is modular) it is possible, at least, in principle to intervene on its variables without directly changing the values of the other variables that affect the *explanans* variable (including the middlemen variables that are between the intervened variable and the *explanans* variable). Simply put, modularity concerns "the separability of causal contributions to an overall effect" (Mitchell 2008: 699).

The aim of the interventionist strategy is to try to guarantee system's modularity and to salvage the separability of causal contributions of different variables whenever the causal system under study allows it i.e when the system actually is sufficiently modular. Therefore, the values of the other variables that are not between the *explanans* and *explanandum* variables and that affect the *explanans* variable should be kept fixed when the *explanans* variable is intervened on (Woodward 2003: 59). In addition, as the condition 2 says, the middlemen variables should not be changed by no means other than changing A: interventions should be so that they guarantee the possibility to separate the causal contribution of the intervened variable A from the causal contributions of the middlemen variables that the variables have independently of A (more on modularity see e.g. Woodward 2003: 48–49, 327 -342).

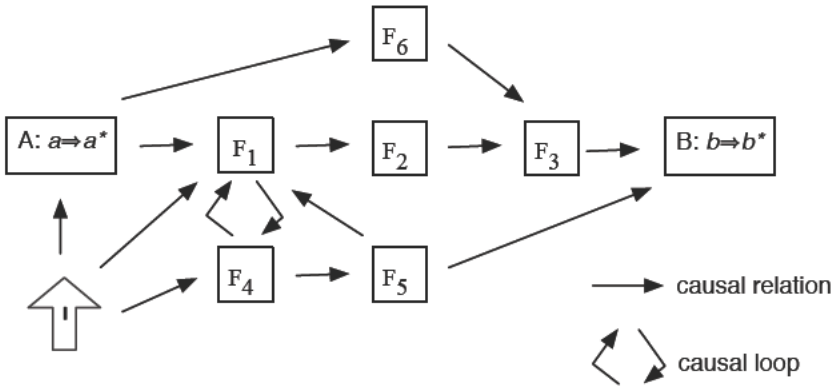
Complexity of a causal system depends on two dimensions: i) the number of middlemen variables between A and B, and ii) the number and nature of the causal relations between the middlemen variables. Complexity of a system is a matter of degree that is influenced by the both of these dimensions. For instance, when there are causal loops between the middlemen variables the system of causal variables is more complex than in the case where the relations are one-way. Causal loops are interactions where the value taken by one variable F_1 causally influences the value of F_2, which then affects the value of F_1 and so on. Complex systems are, therefore, bundles of causal relations and variables that react to interventions in the following way: an ideal intervention on one variable A affects the values of several "middlemen variables" $F_1 \ldots F_n$, and these changes in values might also change the nature of the causal interactions between the variables.

Modularity and complexity can be connected in two ways *in relation to an intervention I on A in regards to B*: there can be *modular complex systems* (picture 5.5) and *non-modular complex systems* (picture 5.6). Both types of systems provide challenges to historians' counterfactual reasoning. In modular complex systems an intervention on A that fulfills the condition 2 is possible, but the intervention on A affects so many middlemen variables that it is hard make out what will be or would have been the effect on B[16]. Inferences on non-modular complex systems are hard not merely because of complexity: they are further complicated by the fact that we cannot intervene on A without directly intervening on some of the middleman variables. This makes it impossible to distinguish the causal effect of A on B from the causal effects of the middleman variables.

[16] This applies to cases like historiography where we are in no position to actually observe the intervention and the subsequent values of B. The change in B caused by the intervention is not observed but inferred on the basis of our evidence.

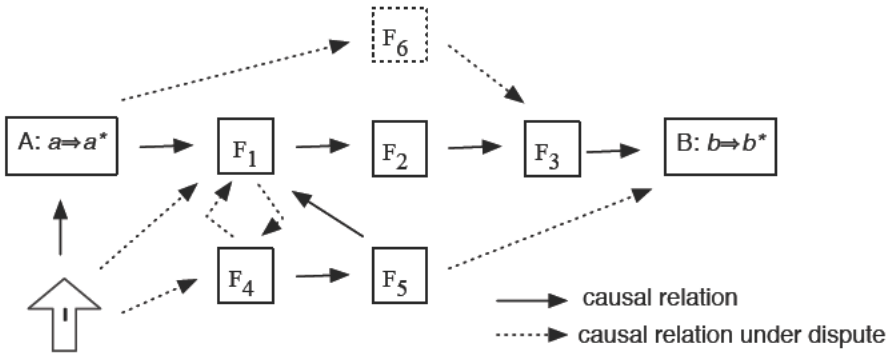


Picture 5.5. *Modular complex systems*



Picture 5.6. *Non-modular complex systems*

Making correct counterfactual what if-inferences on both kinds of complex systems is hard in cases when one actually knows the whole system, the middlemen variables and the principles that govern the interactions between the variables. Historians, however, hardly ever know all the causal relations and middlemen variables in the systems they study. They simply do not have enough evidence to "tell" what the causal system was like. In many cases, however, this kind of detailed knowledge of the system seems to be needed in order to infer what would have happened if A had been intervened. Sometimes it is highly uncertain whether the system that historians study was modular or not. The lack of evidence then further complicates historians' counterfactual inferences and increases their degree of uncertainty. Picture 5.7. is an example of a complex causal system that we know only in part: some of its causal relations and variables are being disputed which means that our counterfactual inferences on the system are equally questionable.



Picture 5.7. Historian's complex causal system

Let me clarify these abstract points with the help of an example. Consider the counterfactual "If the Industrial Revolution had not occurred the British standard of living would have been lower than it was[17]", that J.D Gould (1969) finds deeply problematic. Gould's irritation with this counterfactual is easy to understand.

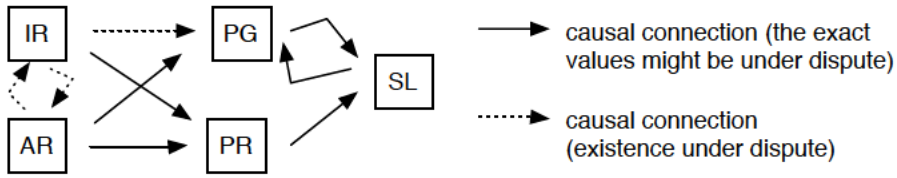
We know that a rise in the industrial productivity increases the standard of living (measured in terms of GDB) in a case where the population stays the same. We also know that population growth affects the standard of living: if productivity stays the same, increase in the population lowers the standard of living. The standard of living reflects back to the population growth[18]. These principles are quite clear, but the actual reality of the process of "the Industrial revolution" was not. When the first signs of the industrial revolution started to appear in the 1770's, Britain's population was already growing because of an increase in agricultural productivity. Economic historians still debate on whether the population growth in the end of the 18th century and beginning of the 19th century was mainly due to the improvements in agricultural productivity or did the industrial developments affect it as well.

On the other hand, it is equally unclear whether the rise of productivity in agriculture was so massive that it would have lead to an improvement in the standard of living, if there had been no industrialization (and the chance to buy grain from abroad with the income earned by exporting industrial products). It is also unclear to what extent the processes of industrialization and the rise in productivity in agriculture where interrelated: if we were to do an intervention on the industrialization process in the

[17] This counterfactual is related to the causal claim: industrial revolution [no industrial revolution] caused the x increase in the standard of living [no x increase in the standard of living].

[18] The models of demographic transition try to make out in what ways the population growth changes as a country industrializes.

1780's and stop the progress in the mining and cotton industries, would this affect the agriculture at all and if it would, in what ways?[19] Because of these uncertainties it is unclear what does the intervention on the industrial process in the 1780's and the sudden end of the progress in the mining and cotton industries would do: how would it affect productivity, population growth and agriculture? Picture 5.8. shows to network of causal relations that economic historians still debate on when trying to evaluate the counterfactual "If the Industrial Revolution had not occurred the British standard of living would have been lower than it was".



IR = "Industrial Revolution"
AR = "Agricultural Revolution
PG = Population growth
PR = Productivity
SL = Standard of living (GDB)

Picture 5.8. Industrial revolution and the standard of living

So what should one do when one is dealing with complex systems and one does not have the perfect knowledge of the causal relations of the system? Many protagonists of counterfactuals recommend that one should not try to do counterfactual inferences where the antecedent and the consequent are far away from each other in time (see e.g. Lebow 2010: 56). This is a reasonable heuristics, since it is usually so that the amount of variables and the interconnections between variables increase when the temporal distance between the antecedent and the consequent increases, which makes it difficult to assess the related counterfactuals. In cases where one has enough evidence on the middlemen variables and their interactions in long-term processes, one is, of course, allowed to start "iffing" on these relations, as well.

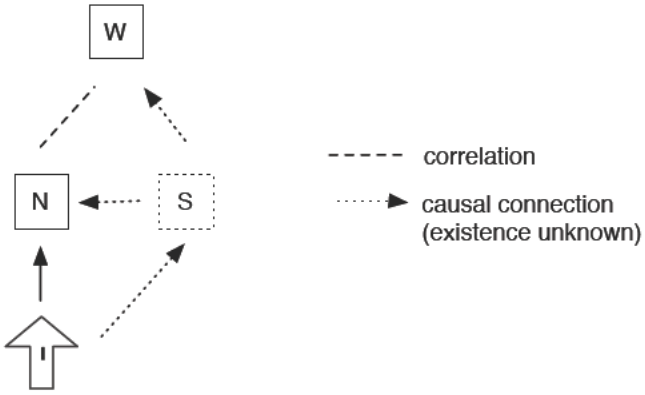
[19] In other words we do not know whether the system is modular or not in this respect: if we cannot intervene on IR without intervening on AR, it is non-modular, if we can, the system is modular in relation to the causal contibutions of these two variables.

5.2.4 Intervention does not correlate with other causes of the *explanandum* variable: the problem of the unknown common causes

The condition 3 says that the intervention on A in regards to B cannot be the cause or even to correlate with some other variable C that is a cause of B. This condition is also based on the need to isolate the causal relation of interest in order to see whether A causes B or not: we are not interested of the overall effect of A and C on B, but only on the purported effect of A to B. For instance, in randomized experiments one needs randomization to exclude the other factors C that are correlated with the intervention and might affect B independently of the causal path between I, A, F and B. For instance, the process that shuffles people into control and test groups in drug trials should not be so that it places the persons with a stronger immune resistance into the other group and the persons with a weaker immune resistance into the other (Woodward 2003: 97). This kind of bias in selection of the test and control groups would lead to a result where the experiment does not test what it is designed to test: the effect of the drug.

Both in quantitative and qualitative comparative case studies one often encounters the problem whether there are unknown factors Z_n that are a common causes of X and Y and therefore account for the observed correlation between X and Y. Only when one is able to rule out common causes for instance with statistical means or process tracing, one can infer that there is a causal relation between X and Y. An intervention on X does not fulfill the condition 3 if the intervention also changes the unknown variable Z that causes systematic changes both in X and Y – based on these on systematic changes between X and Y is possible to falsely conclude that there is a causal relation between X and Y.

Process tracing that seeks to find out the mechanisms between X and Y is not always to pathway to avoid the problem of unknown common causes that are affected by interventions: for instance in his study on the wealth of the Trobriand chiefs, Malinowski is not able to rule out, for instance, the possibility that social status (S) is the common cause that affects both the number of wives (N) and wealth of the chiefs (W). Malinowski is only able to infer that there is some kind of connection between (N) and (W) (Steel 2008: 191–192). Picture 5.9. shows the situation where an intervention affects both (N) and (S). If we are unaware of the variable (S) and of the fact that the intervention on (N) also affects (S), it is in this case possible to conclude that (N) causally affects (W) although it does not.



Picture 5.9. Social status as a common unknown cause

The condition 3, therefore, points out similar kind of problems for historiographical counterfactuals than the condition 2: what is the network of causal relations like in the case of interest? What do we know of it? Are there any external factors C that the intervention is causally connected or correlating with, and if there are, are there some other possible interventions that could be used to eliminate the effect of these external factors C?

5.2.5 Intervention acts a switch: could-inferences and would-inferences

The main message of the condition 4 (*I acts as a switch that controls the change from a to a* irrespective of A's other causes U*) is simple: an intervention should break all the links between A and its causes so that the value of A is completely determined by the intervention. As already pointed out in section 5.2.2, this serves the purpose of isolating the causal relation of interest. Thus, when we do a mental manipulation that seeks to be an ideal intervention, we should not be concerned whether the actual situation was like that that the intervention actually could have realized[20]. As we already saw in section 5.2.2, interventions can sometimes seem like miraculous blows from the sky, because that is what they by definition are. Woodward presents the point as follows (2003: 145):

[20] The reader should note that asking whether a certain intervention was actually possible, plausible, sensible etc. is not the same thing as the question whether a certain cause variable (or its value) is a plausible or sensible cause to account for a certain contrast. The latter question we discussed in the end of section 5.2.2 where the key issue was to separate sensible cause variables from non-sensible ones. The question was not whether an intervention on a certain variable A that would change *a* to *a** actually was possible or not. The considerations on the actual possibility of a certain intervention do not seem to directly affect our considerations on whether a suggested cause is a sensible cause or not.

> It is a striking feature of the kinds of counterfactuals that are relevant to causal and explanatory claims that although we require that they be true when their antecedents are realized by interventions, any more detailed specification of the way their antecedents are realized seems inappropriate and unnecessary.

Historians, however, seem not to agree with Woodward. Historians are immensely interested in questions on under what conditions the antecedents of counterfactuals could have been realized. Historians often present questions of the form "*could* this and this had been different" instead of asking "would this had been different if this and this had been different". *Could*-questions and *could*-inferences seem to be as common as *would*-questions. This is at least the conclusion one must draw when reading through the two of the most important edited volumes that propagate the benefits of counterfactual thinking for historians – *Counterfactual thought experiments in world politics* and *Unmaking the West: "what if" scenarios that rewrite world history*. In practice, almost all of papers in these volumes support the conclusion that Julian Reiss has drawn: "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).

The editors of the volume *Counterfactual thought experiments in world politics* Philip Tetlock and Aaron Belkin (1996: 7 & 18) compress their idea that historians should think about the conditions that are needed for the realization of the antecedent into a *minimal rewrite rule:* one should only present antecedents that require altering as few "well-established" historical facts as possible. The rule can be interpreted so that one should only use antecedents that "do not require us to unwind the past and rewrite long stretches of history" (Tetlock & Belkin 1996: 23). This interpretation reveals historians' way of thinking about counterfactuals in a nice way: Tetlock and Belkin take it as a given that historians "rewrite" history in order to see on what conditions and how easily the antecedent could have been realized. For Tetlock and Belkin it is perfectly acceptable that a historian rewrites the past in order to create story where the antecedent of the causal claim is imagined to have a causal history that guarantees its realization.

Rewriting history means that when a historian imagines an intervention on the variable A and imagines the intervention from *a* to *a** as shown in the picture 5.1. he turns to the causes U of A in order to see whether they would allow the intervention that changes *a* to *a** to actually take place. If it seems that the causes U would not allow the intervention to take place (in the light of U the change from *a* to *a** seems implausible),

the historian considers what changes there needs to be made in U in order to bring about the intervention. Then the historian might even continue moving backwards in the causal chain and look at the causes V of U in order to consider how easily the changes in U could have taken place etc. Thus, historians' counterfactual thinking sometimes works by looking back in the causal chain for the causes of the supposed cause *a*. This causal knowledge is, however, supplemented with extra considerations that are informed by beliefs and assessments on how easy or difficult it could have been to change the causes U of *a*. Together with causal knowledge, the assessments of "easiness" of changes are used to estimate whether *a could actually have been changed to a** in the way the counterfactual says. Many of historians' counterfactuals are *could*-counterfactuals on possibilities: what could actually have happened? The interesting property of them is that they are backward-looking in the way described above.

The contrastive counterfactual theory (the condition 4) states that there is no need to use backward-looking counterfactuals in our causal inferences. The considerations on whether the change from *a* to *a** was actually possible[21] or not should not play a role in choosing explanatory counterfactuals. According to this theory a historian should only be concerned about what happens *after* the intervention on a certain variable in a causal system (the things that happen might show quite a lot of complexity as we saw in section 5.2.3.). The probability of the actual occurring of an intervention does not determine whether the intervention presents an interesting counterfactual claim from the point of view of causal explanation.

Let's illustrate the matter with an example before turning to the reasons why it is always sensible to follow the condition 4 when making causal inferences. Yuen Foong Khong (1996) asks whether it would have made a difference to the beginning of the Second World War if Britain's Prime Minister Neville Chamberlain had acted more resolutely in the face of Hitler's Sudetenland demands. This question concerns the causes of the Second World War: did Neville Chamberlain's actions in Munich play a role in the breaking out of the war, and if they did, what kind of role it was? According to the contrastive counterfactual theory, what a historian should do in answering this question is to try to infer *what would have resulted* if Chamberlain had acted differently in Munich. Inferring what would have resulted does not require that the historian needs to think whether Chamberlain could actually have acted differently. Instead, he needs to

[21] The actual possibility of an event is assessed by means of *could*-inferences, and the assessments are at least in part based on historians' estimations what could easily have happened and what not.

find evidence that tells how the Germans would have reacted if Chamberlain had acted differently and what kind of changes in the field of international politics these reactions would have caused. For instance, documents revealing the strategies that the Germans were visioning in response to the acts of Britain and France would provide this kind of evidence.

Yean Foong Khong presents the explanatory question on the causes of the Second World War in the beginning of his essay and uses less than four pages to review the evidence that helps him to answer this question. The remaining 18 pages he uses to assess whether it *could* have been possible for Chamberlain to act differently, and if Chamberlain could not have acted differently, *could* there have been another person as Prime Minister who could have abandoned the appeasement policy in Munich or before that[22]. Why is Khong like the other authors of the volume so keen on pondering the conditions of realization of their counterfactual antecedents[23]? Does the pondering contribute to the task of causal explanation that the writers take up? If it does, are historian's practice and the theory of causal explanation propagated in this thesis in a clash? Does the clash mean that the theory of causal explanation is utterly wrong in one of its main presuppositions? Or alternatively, are the historians who believe that they are in the quest of finding the causes of events by pondering the plausibility of the antecedent's realization wrong[24]? It is finally time to face these questions that were already implied in section 5.2.2.

---

[22] Asking *could* there have been another person as Prime Minister is not the same thing asking: "what if there had been a another person such as Churchland as Prime Minister. What would have happened then?" The second question concerns the causal role of Neville Chamberlain in the things that happened after the Munich Agreement: was the person of Neville Chamberlain a causal difference-maker for those things that followed or would the things have turned out the same way they did if some other person like Churchland had been the Prime Minister?

[23] I do not believe that the authors of the volume *Counterfactual Thought Experiments in World Politics* ponder the conditions of realization of their counterfactual antecedents merely because the editors of the volume Philip E. Tetlock and Aaron Belkin have asked them to do so (see Tetlock & Belkin 1996:18). Instead, I suggest that there are reasons for presenting the questions on the plausibility or conditions of realization of the antecedent, but these reasons follow from the rationale of causal reasoning.

[24] It is not clear what historians take the relationship between causal explanation and the analysis on how easily the antecedent could have realized to be. For instance, Tetlock & Belkin (1996: 7–9), Fearon (1996: 41) and Lebow (2010: 45) make the remark that historians that work with single cases want to use "plausible or conceivable causes" whereas more theoretically oriented scholars are more easily to accept "miracle causes", but they do not provide any detailed analysis what is the relation of plausible causes or miracle causes to causal explanation. Julian Reiss is more explicit about this: the main claim of his article is that historians' way of thinking about causes differs from philosopher's way of thinking about causes as miracles. In Reiss's view historians have their own method of causal inference that does not follow the counterfactual theories of causal explanation that philosophers (Reiss mentions David Lewis) and some others like the sociologist Max Weber have advanced.

I argue that there are reasons to try to strive to fulfill the condition 4 in one's counterfactual inferences that aim at causal explanations. First, let us go back to the basics: the aim of causal counterfactual inferences is to try to "explore" the causal relation between A and B which in historians case means quite often making sense of the vast network of causal links between A and B. What this means is that the center of attention should be "the stuff" that is between A and B – the way to put the relationship between A and B into focus is to isolate that relationship from other factors, including the actual causal history of A. As already seen, the other conditions 1–3 of ideal interventions also aim at isolating the relation between A and B. The conditions 1–3 are means to pin point the disturbances that non-ideal interventions bring with them: the ideal situation to assess the relation between A and B would be the situation where these disturbances were not present or removed. The point of the condition 4 is similar: if we want to see whether the values of B are dependent on the values of A, then the actual causal history of A seems irrelevant and should, therefore, be "put aside" with an ideal miracle intervention.

Secondly, the idea that interventions work like miracles saves us from the debates on what point of history this intervention could have occurred, could it have occurred at all and so on. For instance, in the case of counterfactual (5.1) where the antecedent described Romania as s superpower, the non-interventionist methodology was to try to rewrite history in a way that Romania would be superpower in the 1970's. Choosing the method of trying to rewrite history so that the antecedent is realized in a plausible way leads to a series of difficult questions. At what time of history should we start this rewrite and on what basis? At the turn of the 20th century? After the Second World War? At the birth of Cleopatra and her nose?

Even if we reached an agreement on the exact point where the rewrite should start, we would be faced with another problem: rewriting even small bits of history, not to talk of decades or centuries, increases the complexity of our causal inferences immensely, and easily we end up in a situation where we are unable to say what would follow from the rewrites. That is why it is better to have one tactical "miracle" or "ideal" intervention: ideal interventions aim at altering as little in the system as possible so that the complexity of the system would not increase immensely. We try to keep the system as simple as possible so that we do not end up at a loss with the complexity of our counterfactual inferences. The Maxim machine gun intervention on the battle of Waterloo in the middle of the battle alters the system so little that we can infer what

would have resulted: Napoleon would have won. Instead, changing the history so that Maxim machine guns would have actually existed, increases the complexity of the networks of causal connections to a point that we are no longer certain whether the whole Battle of Waterloo could have occurred at all. Counterfactual inferences are hard even with "miracle" interventions, and thus, there is no reason to make them even harder with rewrites.

Thirdly, the idea that *could*-questions and *could*-counterfactuals should be used as a criterion in selecting causes for events is misguided. Let us take an example of a simple game with cards. In the game there are cards of four different colors: blue, yellow, green and red. The rules are the following: a blue card can only be followed by a yellow card, the yellow card with a green card and the green card by a red card and red one by a blue one and so on. Let us presume that the game starts with a blue card that is placed on the table. This means that necessarily the third card placed on the table will be green and the fourth one red. Since we know the rules of the game we are justified in saying that the green card on the table caused the placing of the red card on the table and to make the following counterfactual inference: "if the card had not been green, the next card placed on the table would not have been red".

The fact we know that it was necessary that the third card was green since the game started with a blue card (and that it was impossible for the third card to be of any other color) does not prevent us from making the above-mentioned causal and counterfactual inferences. Why not? That is because our causal judgments are informed by knowing the rules of the game, not our knowledge on what was necessary, possible or impossible given the causal history of the system. Knowing the rules allows to imagine interventions that place a blue, red or yellow card in place of the green one and to infer that in these cases the next card would not have been red. This counterfactual inference helps us to conclude that it was the green card that was the cause of the placing of the red card. The fact that it was not "objectively" or "actually" possible that the green card could have been of different color does not prevent us from making the causal claim that the green card was the cause of the red card.

In the same way, the knowledge that Neville Chamberlain's hands were politically tied and that it might have been impossible for him to act differently in Munich and to abandon the appeasement policy, does not prevent us from making causal and counterfactual inferences on the consequences of the appeasement policy. If we have evidence that the Nazis would have changed their military strategy if

Chamberlain had been tougher in Munich, we have a reason to think that Chamberlain's behavior causally influenced the beginning of the Second World War. This evidence gives as a reason to state the counterfactual antecedent "If Chamberlain had acted differently". The antecedent makes sense as a part of an explanatory causal claim independently of the fact whether Chamberlain could actually have acted differently or not. Knowledge on what was actually possible and *could have happened* does not affect our causal inferences on *what would have happened.* The questions on what could have happened and what would have needed to be different so that a certain antecedent could have realized (*could-counterfactuals*) should be kept separate from the causal questions on what would have happened if this had happened (*would-counterfactuals*).

Why are *could*-counterfactuals so popular among historians? Could-*inferences* are used to answer two kinds of questions that seem to interest historians and general public alike. First of all, there is the question of prevention: could a certain event or an outcome that we wish not to have happened been prevented? Could the First World War been easily prevented? Can we ensure that this thing does not happen again? What kind of things we need to change in order to prevent these kinds of things from happening again? To answer these questions we need to have an account of the causes of the event, but in addition to the causal knowledge provided by *would*-inferences, *could*-inferences are used to estimate how easily these causes could have been changed so that the unwanted outcome would not occurred.

The questions of prevention lead quite often to questions of responsibility, guilt, praise and blame[25]. The questions on whether certain actors could have done to something to prevent the bad things from happening are often raised. Was Neville Chamberlain in a position to break free from the appeasement policy and even to prevent the Second World War from breaking out? If Neville Chamberlain was in a position where he actually could not break away from the appeasement policy, he is not to be blamed for following that policy as much as he would be if he had been totally free to choose his strategy in Munich. Since the blaming and praising of historical actors is a common phenomenon in the public discussion in many modern societies, historians can contribute to this discussion by trying to deliver to the public the best

[25] The reader might remember from the chapter 2 that also Daniel Nolan (2001: 15-17) took notice that historians use counterfactuals to deal with the questions of moral responsibility. The theme of moral responsibility was also discussed in the chapter four where it was argued that a historian should try to keep the roles of researcher interested in causal explanations separate from role of the judge. This section has showed that playing the role of the judge by using *could*-counterfactuals might be confusing from the point of view of causal explanation.

historiographical knowledge that exists on the conditions that prevailed when an actor did what he did. Knowledge on the conditions that prevailed helps to answer questions whether the actor could have acted differently.

Like in the case of prevention, raising the finger of praise or blame often requires that one has causal ideas that the actions of the person to be blamed or praised actually contributed to the occurrence of a result that is considered either good or bad (like the beginning of the Second World War). Therefore, blaming or praising Neville Chamberlain for the appeasement policy would makes sense only when we are convinced that the appeasement policy played a role in the beginning of the Second World War[26]. So before historians get mixed with the issues of moral blame and praise they should have at least some kind of knowledge of the causal relations that prevailed[27].

Claims that a certain person was the only person who was able to bring about a certain result (that we attribute a moral value to) provide another interesting case of attribution of moral praise. For instance, there are arguments that the actions of the Athenian military leader Themistocles not only led to the winning of the Battle of Salamis 480 BC, but that he was *the only* Greek military leader to be able to plan the victorious tactics against the Persians (Hanson 2006). If Themistocles' tactics lead to the Greeks' victory against the Persians, he is praised. If Themistocles was the only one cunning enough to design the tactics that he did, he is praised even more, since this means that no one else would have been able to save the Greek culture and the seeds of Western culture – a result that the modern westerners often value.

The arguments that a certain result would have not happened without certain acts or events are ubiquitous in historiography. What kind of counterfactual thinking are these kinds of claims based on? On what basis can one make the argument that Themistocles was *the only* one who could have won the Battle of Salamis? This is the topic that I am going to return to in the next chapter where I discuss the various uses of interventionist counterfactuals.

---

[26] When people blame or praise historical actors, usually some kind of consequentialism prevails: actions are deemed good or bad based on their consequences.

[27] The *could*-question on possibilities and the *would*-question on causal dependence are usually presented together. For instance, in economic history there has been a huge debate on the questions whether the recent African history could have taken another course than the actual one, and what would the consequences have been for the modernization and the economic development of the continent (Elster 1978: 192–196). The way these questions are often presented and investigated together probably makes it difficult for practicing historians to keep the related counterfactuals separate: *could*-questions require backward-looking counterfactual inferences, whereas *would*-questions do not.

6 Uses of interventionist counterfactuals

Is it legitimate for the historian to consider alternative possibilities to events which have happened?... To say that the thing happened the way it did, is not at all illuminating. We can understand the significance of what did happen only if we contrast it with what might have happened.

– Morris Raphael Cohen

In the previous chapter we saw that historians employ two kinds of counterfactual inferences: firstly, they use backward-looking counterfactuals to infer things that could have happened and secondly, they use (or as I claim they should use) interventionist miracle counterfactuals to infer causal dependency relations. In this chapter, I will discuss the most important ways in which the interventionist counterfactuals are put to use in historians' causal reasoning. I will start with separating four forms of of causal weighting, and to conclude, I will briefly discuss how causal reasoning and counterfactual inferences work in the construction of historiographical narratives.

6.1 Causal weighting

The metaphors of "causal weighting" and "weighting of causes" reveal one crucial role that counterfactuals play in historian's causal reasoning. E.H. Carr (1964: 90) once claimed: "Every historical argument revolves around the question of priority of causes". Historians are not merely interested in finding the causes of events, but they are also keen on comparing the suggested causes in order to assign relative importance to them: was this factor more important than the other causes in bringing about the effect or not? This section discusses the various ways in which historians "weight causes" with the help of counterfactual inferences in order to reveal their "significance", "importance" or "meaning" for the *explanans*[1].

[1] Historians talk very often of "significance", "importance" and "meaning" of causes, but as Martin (1989: 53) notes philosophers of historiography have not paid too much attention to the question of causal weighting, although the theme has been raised many times in the context of counterfactual thinking. Martin , as well as de May & Weber (2003) and Robert Northcott (2008), all seek to solve the problem of explanatory weighting with the help of contrastive *explananda*. Maw Weber (1949) discusses the attributions of causal significance and "adequate causation" based on the considerations of necessity and contingency that are reached with counterfactual inferences. None of the analyses that I am aware of analyses the phenomenon of causal weighting as composing of several ways of inferring with the help of interventionist counterfactuals, as it is presented in this section. It seems that the presupposition has been that "causal weighting" is a monolithic activity composing of one type of questions and one type of counterfactual inference.

6.1.1 Selecting causes and fixing the causal background

The first way to employ interventionist counterfactuals to do causal weighting is to separate causes from the background factors. This is the question that was discussed in the two previous chapters. In chapter four it was pointed out that presenting well-specified contrastive explanatory questions helps one to determine which factors to include in one's explanations as causes accounting for the contrast and what to leave out as background conditions. This means that the factors that do not answer the specific contrastive *explanandum* should be treated as background factors. The examples of why Kennedy won instead of Nixon and why the U.S. election system is the way it is showed clearly how the causes can be separated from the background factors. Background factors explain different contrastive questions than causes. In other words, background factors are the factors that do not make the difference presented in the contrastive *explanandum.*

In chapter five we saw that among the possible causes that explain a certain contrasts, some causes are believed to be more sensible than others. The considerations on what to count as sensible causes and what not can also be said to be a way of separating background factors from causes. The causes that we judge as insensible can be treated as background factors: there is no reason to cite the lack of modern machine guns or meteorite strikes as causes for the defeat at Waterloo, although these factors do, in fact, account for the difference presented in the contrastive *explanandum.*

In addition, chapter five dealt with the difficulties that historians face in their causal inferences when determining causes. The most important problems of causal and counterfactual inferences in historiography stem from two sources. First of all, historians often lack evidence that could support counterfactual and causal inferences. Secondly, historians often deal with complex causal systems where there are so many interacting variables and where also the "rules", according to which the variables interact, can change when the system is intervened on. This means that it is hard to do "surgical manipulations" on the system and to change only the values of the variable of interest, which would make counterfactual inferring neat and simple. Historians do not often have enough evidence to infer what the non-ideal manipulations that they are able to perform would do in the complex causal system of interest: would the intervention that miraculously changes *a* to *a** bring about the change in B, and if so, what kind of change?

6.1.2 Using various contrastive *explananda*

Let us leave these issues of separating the background from the causes and problems of inferring the right causes behind us, and assume a situation where a historian has plenty of hypotheses on the causes of a certain outcome. This is, in fact, quite often the case. Historians have plenty of causes that they believe to have contributed to bringing about a certain outcome. The problem then is how to do causal weighting among these. In other words, historians are not merely interested in knowing whether the facts *a*, *c*, *d* and *e* were together sufficient to bring about the value *b* in B. They want to ask more specific *what if*-questions on the causal roles that *a*, c, *d* and *e* played in bringing about *b*: if we change *c* to *c**, what kind of change will it bring about in B. Is it so that that no event like *b* occurs, or does *b* occur altered into *b**? This means presenting more specific contrastive *explananda* than just asking why *b* occurred instead of not occurring. Historians want to know whether all the factors *a, c, d* and *e* accounted for the specific timing of or to a specific feature of *b* or was it so that only some of these factors accounted for the various nuances in *b*.

Presenting various contrastive questions of a certain outcome is, therefore, one way of doing weighing among causes. It is a way of finding out what specific aspects of the outcome certain causes accounted for. The example of the Industrial revolution presented in chapter four is an example of the use of this kind of methodology of presenting various *explananda* of a certain outcome In addition to the method of seeking causes to various aspects of the same outcome there are two even more specific means of doing causal weighting: the analysis of contingency and the analysis of sensitivity.

6.1.3 Analyses on contingency and overdetermination

In the end of the fifth chapter, I introduced the following claim (6.1): Themistocles was the only Greek leader that was able to beat the Persians in the Battle of Salamis. What kind of counterfactual claim is this and how does it differ from the claim (6.2): Themistocles caused the Greeks' victory? Hanson (2006) who discusses the first claim (6.1) supports it with the following evidence: there is evidence that all the other Greek leaders of that time would not have made the similar kind of plan that Themistocles did. The other leaders would have prepared tactics that would have likely resulted in a loss. This evidence provides support for the counterfactual: "if Themistocles had been

replaced (by a miracle intervention) by some other Greek military leader of the time, the Greeks would have lost the Battle of Salamis".

The causal claim (6.2) that it was Themistocles (or to be more precise his tactical plan) that caused the victory is, on the other hand, supported by the evidence on the conditions that prevailed that day, detailed knowledge of the Persian and Greek fleets, their sizes and normal tactics. This evidence helps to conclude that it was the way that Themistocles decided to use the Greek fleet that caused the victory. For instance, if the Battle had been fought at open sea instead of in the narrows of the channel of Salamis, the Persians would have lost.

The claim (6.1) that Themistocles was the only Greek leader who was able to win presupposes the causal claim (6.2) and the related counterfactual inferences. Before one can say that Themistocles was the only man able to win, one needs to have evidence that it was Themistocles in the first place who caused the victory. The counterfactuals that support the claim (6.1) that it was Themistocles alone who was able to win the Persians are therefore different from the counterfactuals that are used to define causes. Causal inference makes use of interventionist counterfactuals that "test" on the basis of evidence whether there is a right kind of dependency relation between *a* and *b*. Claims like (6.1) are, on the other hand, used to raise the question *whether there were other causal routes that would have realized if the fact [a] had been removed altogether and whether this would have led to an outcome that would have been somewhat similar to [b]*. Answering this kind of question requires evidence on two things. Firstly, one needs to have evidence that there were other causal routes that would have realized if *a* was removed. Secondly, one needs to know what would have resulted from the realization of these other routes. In other words, one wants to know whether the actual route of where *a*[*a**] caused *b*[*b**], was the only route that could have led to an outcome similar to *b*[2].

When one removes the actual route of causes and analyzes the alternative routes of causes that would have resulted from removing the actual route and tries to define whether the outcomes of these routes would have been similar or not to *b*, one analyzes the degree of *contingency* of the outcome *b*. If there is evidence suggesting that removing the actual chain of causes was not likely to lead to developments that would have resulted in an outcome similar to *b*, *b* is judged as *contingent*. If it seems on the

[2] Philosophically, this kind of inference leads to difficult questions on the identity of facts or events: on the basis of which criteria something is judged to be similar to *b*? But since historians seem to operate with our unexplicated notions of similarity when answering these questions, I won't get mixed with the issue what is really means to say that *x* is similar to *y*.

basis of evidence that the removal of the actual chain of causes had led to developments that would have resulted in an outcome similar to *b*, *b* is said to have been *necessary* or *overdetermined*. It is important to notice that this kind of empirical analysis on the degree of contingency of an outcome is far removed from the metaphysical questions of necessity and contingency to which some antagonists of counterfactuals claim counterfactual thinking to lead.

In order to illustrate this type of analysis a bit more, let's discuss the breaking out of the First World War – a theme that has attracted a lot of counterfactual "iffing" that has aimed at revealing whether the war was necessary or not[3]. The discussion has revolved around the question: if there had been no Sarajevo shots, would a war similar to the First World War have broken out sometime after July 1914 or in the years after 1914? The idea of this question is to weigh the importance of two factors that we believe to have contributed to the breaking out of the First World War. We believe that the alliances that the European superpowers formed in the late 19th and early 20th centuries played an important role in the breaking out of the war: without this kind of alliance structure the Sarajevo shots, which we regard as the other important cause of the war, would not caused a turmoil that dragged so many European countries to war. So, no war due to the Sarajevo shots without the alliances. This is the inference that most historians can agree on. What is the source of much more debate is the question whether there would have been a war without the Sarajevo shots. In other words, among the actual causes of the war, were the alliances more important than the Sarajevo shots in bringing about the war? To answer this question we have to find evidence that could be used to conclude what would have happened if the Sarajevo shots had not been shot or had been removed by a miracle intervention: are there reasons to believe that removing the shots had resulted in alternative causal routes that would have lead to a similar kind of war?

How does this kind of analysis of contingency and overdetermination relate to the classic question of overdetermination in philosophy? Overdetermined causal systems have posed a vast challenge for philosophers defending counterfactual theories of causation, since simple counterfactuals of the form "if *a* had not occurred, *b* would not have occurred" fail to capture the causal relations in overdetermined systems. In cases of overdetermination there are two (or more) causes, *a* and *d*, present that both

[3] For a nice summary of the discussion, see Lebow 2010: ch. 3.

will lead to the occurrence of *b*[*b**] independently of each other. The counterfactual "if *a* had not occurred *b* would not have occurred" is false since if *a* had not occurred, *d* would have still caused *b*[*b**]. What is a troubling technical worry for a philosopher[4], is a remarkable empirical question for a historian, however. If a historian manages to find evidence for the claim that the beginning of the First World War was overdetermined, she is entitled to make various inferences on the beginning of the war. She can, for instance, claim that the alliances were a more important causal factor for the beginning of the war than the Sarajevo shots, since even without the shots the alliances formed an explosive cocktail that was ready to explode if only it was given a spark like the Sarajevo shots. She can also conclude that the shots can be used to answer the question why the war broke out on July 1914 and not some other time, but that the shots do not provide a sufficient answer to the question: why did the war break out instead of not breaking out at all? In order to account for the latter contrastive questions she needs to cite other causes[5].

The analyses on contingency are a form of causal weighting where a historian tries to determine with the help of interventionist counterfactuals what factors in the causal chain of events leading to an outcome were more important than others. Was the context or certain factors of the chain of events causally significant for a certain result so that removing a part of the causal chain or the whole chain[6] would not have mattered for the outcome?[7] The analysis of contingency is about removing the actual chain of causes and seeing where the interactions between other factors (the background factors)

[4] Woodward (2003: 77–79) claims to have solved the problem of overdetermination with his interventionist account. Interventionist counterfactuals manage to capture the right causal structure in cases of overdetermination: if *b* is overdetermined by *a* and *d* and if *a* is the actual cause of *b*, interventionist counterfactuals lead to the result that it is *a* that causes *b*, not *d*. Simple non-interventionist counterfactuals, on the other hand, would lead to the wrong conclusion: *a* is not the cause of *b*.

[5] If on the other hand, the historian would find evidence to conclude that the Sarajevo shots were the only causal route that leading to war, she is entitled to attach much more causal importance to the shots: the shots not only explain the timing of the war, but also are an important factor in explaining why the war broke out in the first place. In this case the beginning of the war would seem highly contingent.

[6] It depends on the case under study whether is it more feasible to talk of removing a part of the chain rather than removing the whole chain. In the case of Themistocles and the Battle of Salamis it is more natural to talk of removing the whole causal chain, since there the intervention on the leader of the battle means changing the whole course of the battle. In the case of the First World War, removing the Sarajevo shots and the things that followed the shots is more naturally described as removing one part of the causal chain that lead to the First World War.

[7] Analyses on contingency and overdetermination are also ways to determine causal *turning points* in history. Turning points are factors that are difference-makers to various events that succeeded them: if we remove or change the turning point in some way, it would result into massive changes. For instance, it has been clamed that if the Greeks had lost the battle of Salamis, the Western culture had not developed at all (Hanson 2006; Strauss 2006). The birth and rise of the Western culture are thus seen as highly contingent: there is one single event, the victory of the Battle of Salamis that the Western culture was depended on. If this event had been removed, the Western culture would not have developed.

would have led after the removal. The fourth form of causal weighting – sensitivity analysis – that I will discuss next is the mirror image of this kind of inference: the point of sensitivity analysis is to change (or remove) some of the other factors in order to see whether the actual chain of causation would have stayed the same.

6.1.4 Analyses on sensitivity

Counterfactual dependence and invariance are not the only significant properties that causal relations have. Causal relations and the related counterfactuals also have the feature that can be called 'sensitivity' (Woodward 2006: 2; see also: Ylikoski & Kuorikoski 2010). A causal claim and the related counterfactuals are sensitive if they hold in the actual circumstances, but fail to hold in circumstances that depart in various ways from the actual circumstances (Woodward 2006: 2). To the extent to which a causal claim and the related counterfactual would fail to hold under various changes in the actual circumstances, defines how *sensitive* or *insensitive* the causal claim and the related counterfactuals are.

For instance, pushing the break when driving a car makes it stop. This causal claim is insensitive to various changes in the background conditions. For instance, the speed of driving does not affect the whether the car eventually stops or not (it, of course, affects the time it takes for the car to stop). It does not matter whether I am driving in rain, on a German highway or with my high heel shoes on for the car to stop or not. The causal relation is, on the other hand, sensitive to changes in the break system of the car: if oil gets into the break system, the car might fail to stop.

Sensitivity analysis is a form of causal reasoning that employs interventions and counterfactual inferences: one tries to find evidence to support one's inferences on what would happen to the causal relation of interest under various interventionist changes in the circumstances that actually occurred. Successful sensitivity analysis contributes to our causal knowledge by further expanding the area of *what if*-inferences we can make: it gives us the ability to answer correctly not only *what if*-questions on the invariance of the causal relation but to *what if*-questions on the holding of this invariant relation in changing circumstances. Sensitivity analysis can, therefore, be seen as a form of causal weighting: how important was the fact that the other factors took the values they actually took for the causal relation to hold and to produce the outcome it did? In other

words, how did the actual circumstances causally contribute to the fact that the causes actually were able to produce the effect?[8]

In historiography, there are several examples that can be interpreted as sensitivity analysis. To further expand Hanson's (2006) case study on the Battle of Salamis let us bring up the following causal questions: What was the importance of the weather conditions being the way they were or the Persians ships being the way the were or the color of Themistocles shoes for the holding of the causal relationship that brought about the victory of the Greeks? Would Themistocles' plan have worked if the color of his sandals had been different? Would it have worked if a storm had struck in the middle of the battle or if the Persian boats had been of similar size to Greek boats? What if the "war hawk" governor-general of Bosnia-Herzegovina, Oskar Potiorek, had been killed in Sarajevo in 1914 instead of the more peace-loving idealist Franz Ferdinand? Would this have made a difference in the causal process that lead to the First World War (see Lebow 2010: 83–84)?

What can historians achieve with these kinds of questions and sensitivity analyses? First of all, knowing what kind of factors the holding of the actual causal relation was dependent on helps us to make informed decisions on what kind of factors to ignore as irrelevant background factors to be excluded from the historians' descriptions of causal processes. If we are entitled to infer that an intervention on the color of Themistocles' shoes right before or in the middle of the Battle of Salamis, does not make a difference to the carrying out or succeeding of Themistocles' plan, we have a good reason not to include Themistocles' shoes into our descriptions on the causal processes on the victory. The weather conditions and the size of the Persian boats, on the other hand, seem to be background factors that needed to be quite close to their actual values in order for Themistocles' plan to work. Therefore, sensitivity analysis helps to determine the factors that were irrelevant to the success of Themistocles' plan and the factors contributed to the success of Themistocles' plan. This way our picture of the causal processes governing the Greek victory becomes "thicker" and fuller: we do

[8] Sensitivity analysis can also be regarded as a way to make inferences on the contingency or necessity of a result (see Ben Menahem 2011 for further discussion on this). If the causal relation between *a* and *b* is extremely sensitive, we usually consider *b* to be highly contingent. If, on the other hand, the causal relation is very insensitive to changes in the background conditions, *b* is regarded as a necessary result: we would have needed a lot more interventions (or one "massive" intervention) on the background factors in order to stop *b* from occurring. The analysis on contingency and necessity of events that makes use of sensitivity analysis is not the same thing as the analysis of contingency and overdetermination presented in the previous section. Sensitivity analysis concerns changing the background factors and analysis of overdetermination works by changing or removing part of the causal chain of interest.

not merely causally understand the causal relation, but also the relevant background conditions that explain why the causal relation held. This way we also get information on what factors one should have intervened on in order to prevent the causal process and the outcome of the process from occurring – interesting information for historians who are also concerned about questions on how a certain unpleasant result could have been avoided[9]?

It seems that many causal relations that historians deal with are relatively sensitive: there are usually quite many interventions on the actual values of several variables that could have affected the causal relation. This kind of sensitivity of the causal relation also blurs the identity of the causal relation itself: should we count the dozens or hundreds of "background conditions" that do make a difference as endogenous parts of the causal relation or not? I do not have any general recipe to solve this question: the issue needs to be settled case-by-case and sometimes more detailed contrastive questions might help in cutting down the abundance of causally relevant factors.

As counterfactual inferences on invariance, the analyses of sensitivity sometimes turn out to be extremely challenging. What makes counterfactual inferences on sensitivity sometimes difficult are, however, the same problems of complexity and modularity that make counterfactuals on invariance difficult. If the network of causal relations that is of interest to a historian is complex in the way that it involves dozens of variables most of which interact with each other, counterfactual inference becomes difficult. For instance, should we interpret the intervention on the weather during the Battle of the Salamis in the way that this intervention would not have affected the carrying out of Themistocles' plan? Are there reasons to assume that Themistocle's would have reacted to a sudden change in weather by revising his plan altogether? If Themistocles had revised his plan, what would have resulted? Do we have enough evidence to know that? The biggest problem for historian's causal inference is not whether a certain intervention could have realistically occurred or not. The biggest

[9] For instance in his discussion of the First World War, Richard Ned Lebow asks whether the First World War *could* have been avoided or not? This question is intriguing for a historian as an intellectual challenge, but also important for the general audience and politicians concerned with the issues of responsibility. In order to answer this question one needs, first of all, to determine with the help of interventionist counterfactuals the factors that the holding causal relation was depended on. After that one employs *could*-counterfactuals to determine how easily the factors on which the holding of the causal relation was dependent could have been changed.

problem is that historians sometimes do not have enough evidence to infer what would have happened in the network of causal relations after an intervention.

The problem of defining what would actually happen after an intervention is not a problem that only historians face. James Bogen (2004: 5–6; 2004: 5–6; see also Hausman and Woodward 1999: 537, n.12) introduces an example from astronomy:

> Suppose the mutual gravitational attraction of the heavenly bodies makes it impossible for anything to change the moon's distance from the earth enough to increase or to diminish its influence on the tides without changing the positions of other heavenly bodies so that their influence on the tides changes significantly. Then if something moved the moon closer or farther from the earth, the ensuing value of the tides would reflect the causal influence the other heavenly bodies exert independently of the moon's influence on the tides and moving the moon would not qualify as an intervention on the moon's distance from the earth with respect to its tides.
>
> (Bogen 2004: 5–6.)

Ideal interventions are hard to make in systems that have multiple interacting parts, since changing one part would affect other parts in a way that it becomes hard to estimate the end result. The contrastive counterfactual theory does not help to solve the problem of how to deal with complex systems where an intervention affects not just the variable, but several variables through the interactions between these variables. It seems to be a task that is left to the practicing scientists and historians to figure out what kind of and how much of causal knowledge can be gained from the actual causal systems that do allow clean-cut surgical interventions. However, this does not render the contrastive counterfactual theory useless. The theory still provides guidelines where to move in our causal inferences: what kind of questions to ask and what kind of experiments to design and what kind of evidence to look for and use to support causal claims. As seen in the chapters in this thesis, this is quite a contribution. Before closing this chapter let us take a look at one more implication of this theory and explore the relationship between counterfactual inferences and historians' main representative means – the narrative.

6.2 Explaining with narratives

As the previous sections have shown historians usually need to deal with a variety of causes when constructing causal explanations. E.H Carr makes a point by claiming (1964: 89): "[t]he first characteristic of the historian's approach to the problem of cause is that he will commonly assign several causes to the same event." Dealing with several

causes leads to the problems of causal weighting described in the previous sections, but it also creates another problem: how to present an abundance of causes in a cognitively salient way? Historians have solved this problem by making use of the distinct human capacity to tell stories, and thus narratives have become the main tool of representation in historiography.

Originally historiographical narratives resembled stories in many respects: they presented in chronological order how certain events interacted together to produce a certain outcome. In current historiography, narratives have become more versatile: the causes of events and consequences are not necessarily presented in chronological order (Fellman & Rahikainen 2012: 9–11). Instead, more thematical approaches are preferred. The causes might be presented grouped, for example, in structural and individual level causes, and what is more important, chronological order is often broken with considerations of the significance or importance of causes.[10] Despite these changes, the central aim of the narrative style of representation has stayed the same: the aim is to causally explain *how* the event of interest came about[11]. The last section of the thesis deals with the question how counterfactual inference contributes to narrativist causal explanation.

6.2.1 Counterfactual inference and the construction of narratives

There has been a long-standing debate in philosophy of history from the 1960's onwards on the question whether "narrative explanation" is an independent form of explanation and whether narratives explain in virtue of being "narratives" (see e.g. Dray 1971; Carr 2008; Froeyman 2009). The claim of this section is that narrative explanation is not an independent form of explanation and that narratives are not self-explanatory. The question is therefore the following: when a historian works with an abundance of suggested causes in order to construct a narrative, how does counterfactual reasoning work and how the reasoning contribute to the fact that narratives explain?

Let us start from the problem of explanatory selection that was already touched upon in the chapter four in section 4.2 that discussed the role of contrastive *explananda* in selecting explanatory causes. The issue can be further illustrated with another

[10] Sometimes the considerations of significance are presented by means of explicitly presented counterfactuals.
[11] Causal explanatory questions can also have the form of "how"-questions: how did *b* come about instead of *b**?

problem of selection: table 6.1 shows some of the factors that have been presented as the "underlying" causes for the First World War. Let us imagine a historian who wants to write a narrative representation of how the First World War came about. How does she choose from the suggested causes the ones that she presents as causes and how does she present the interactions of these causes?

DEEP CAUSES	**Industrial revolution**: war as escape from super ego constraints) **Modernization**: war as *raison d'être* for traditional classes and values **Nationalism**: threatened survival of multinational empires **Democratization**: threatened traditional political elites
SYSTEM LEVEL	**Imperialism:** German need to acquire more colonies **International anarchy:** Struggle for relative gain, arms races **Alliance structure:** alliances too tight; alliance too weak; need to preserve Austro-German alliance **Shifts in the political balance:** encirclement of Austria, Russian fear of losing great power status **Shifts in military balance:** Russia growing stronger and Austria weaker **Changes in the rules of the game:** imperial competition in Europe
IDEAS	**Social Darwinism:** international politics as struggle for survival **War is inevitable:** German fear of Franco-Russian intentions **War is beneficial:** expected to temper the nation **Code of honor:** required war in response to certain provocations
STATE STRUCTURE	**Weak society and strong state:** domestic conflict is exported
DOMESTIC POLITICS	**Conservative desires to safeguard status quo:** German conservatives sought war to crush socialism; Austrian fear of Slavic nationalism: Russian fear of revolution **Weak regimes**: Serbian sponsorship of Pan-Slavism
ORGANIZATION AND BUREACRACY	**Military planning in political vacuum:** the Schlieffen plan, Russian failure to develop fully separate mobilization plans against Austria and Germany **German military pressure for preemptive war** **Cult of the offensive:** preference for offensive over defensive strategies **The short war illusion**: victory was possible at low cost to society **Rigid organizational routines:** restricted freedom of political leaders
LEADERS	**Poor German leadership after Bismarck**: led to German isolation and encirclement **Poor German crisis management:** misread military balance; misread responses of other actors; fear and panic interfered with judgment **Poor Austrian crisis management:** misread military balance and likely political consequences of war with Serbia; failed to convince Russia of Serbian complicity in the assassination; Conrad's purely personal motive for war **Poor Russian crisis management:** failure to recognize consequences of partial mobilization **Poor British crisis management:** belated recognition of Austro-German plans by Grey

Table 6.1. Some of the causes presented for the First World War (source: Lebow 2010: 75–76)

Obviously, the historian needs to do vast deals of archival work in order to make sure that the hypothesized causes actually existed: did there actually exist a Darwinist idea of the survival of the fittest among the leaders of European countries at that time and did the idea contribute to their decisions in the events that followed the Sarajevo shots? Archival work also gives the historian hints about the ways in which the various

factors came together in the process that led to war. In the end, however, counterfactual inferences play a part in the selection of causes and explanatory factors. The historian weighs causes in order to decide what factors are more important than others and these considerations play a role in what the historian leaves out and how she presents the factors that she includes in the narrative. The central claim of this thesis has been that the various forms of causal thinking that historians engage in are also forms of counterfactual thinking. The selection of causes, analyses of contingency and sensitivity all make use of interventionist counterfactuals and all these can play a role in the construction of narratives.

In the light of the contrastive counterfactual theory, narratives depict chains of causation that are believed to have resulted in the outcome(s) of interest. In other words, narratives aim at depicting the factors that are considered on the basis of counterfactual inferences to have been the difference makers to a certain outcome. This is the feature that separates narratives from mere chronicles that present events followed by other events. Chronicles are *descriptions* of the form "this event was followed by another and after that something else happened". Narratives, instead, are *causal accounts* – usually full of causal verbs like *led to* or *fostered* – that aim at depicting chains and bundles of causal relations that answer the contrastive question: Why this outcome instead of some other kind of outcome? Among the causes that answer this contrastive question one can also do further causal weighing. For instance, if one concludes that the conservative desires to safeguard status quo in Austria, the belief that the war is inevitable among the German military officers, as well as poor Russian crisis management (see table 6.1) were all causes that played a role in the beginning of the First World War, one can ask further: What kind of role? What kind of more specific contrastive questions can these factors answer? Are some of these factors more important than others in explaining a certain contrasts? These are all forms of causal reasoning that affect the construction of narratives. The conclusion of these considerations is that *narratives do not explain in virtue of being narratives, they explain in virtue of representing complex chains and bundles of causal relations that are believed to support various counterfactual what if-inferences.*

However, the narrative form usually hides the *what if*-inferences that affect the construction of narratives. The narratives themselves contain explicit counterfactual claims only occasionally: their explicit use is usually confined to situations where a historian discusses the causes she considers to be most significant ones. Hence,

counterfactual reasoning remains in many cases implicit, and it might be that the historian does not pay too much attention to the inferences that guide her causal thinking: causal and counterfactual thinking stays hidden behind the phrase “historian’s imagination”.

There are dangers in this implicitness, however. For instance, implicit counterfactual thinking might lead to situations where historians think that they have more causal knowledge than they actually have: a coherent story of the way things evolved might give us the impression that we know the causes of a certain event. Asking counterfactual “what if”-questions might reveal this is not the case. When one approaches evidence with the help of counterfactual questions, it might reveal that the evidence is all but conclusive in determining what to count as causes and what not. The explicit use of counterfactuals might prevent one from doing hasty causal conclusions, and also give one strong reasons to believe in one’s causal conclusion: if one can present the evidence that supports one’s counterfactual claims, the evidence is evidence also for the related causal claim.

Therefore, there are better guides than the mere unarticulated power of imagination to direct the construction of narratives. The explicit use of counterfactuals and contrastive explanatory questions and various forms of counterfactual thinking contribute to the quality of historiographical research – they force one to think more explicitly the content of one’s explanatory causal claims and the evidence that one has to support these claims. Therefore, I hope that, in the future, the narrative style will also witness an increase in the explicit use of counterfactual inferences as well as an increase in the evaluation of evidence in the light of these inferences.

7 Conclusions

There is no a priori guarantee that every subject matter must be a suitable domain for causal explanation.

– James Woodward (2003: 314)

The task set for this thesis in the end of the second chapter was to introduce a theory of causal explanation that would fulfill two desiderata: firstly, the theory should explicate the link between causal explanation and counterfactuals in order to assess the claims of both the antagonists and the protagonists of counterfactuals. In other words the aim is to provide a balanced solution to the debate between the two sides. Secondly, the theory should not only explicate the relation between causal explanation and counterfactuals, but also do it in a way that can be shown to be methodologically fruitful. These are the two desiderata set from the side of philosophy of historiography and practicing historians. From the point of view of philosophy of science the most important task is to investigate whether the most distinguished theory of causal explanation of today applies to historiography, and if it does how well? Now its time to summarize to what extent the theory of causal explanation presented in this thesis – the contrastive counterfactual theory – copes in fulfilling all these three tasks.

For the protagonists of counterfactuals the conclusions of the thesis are soothing: there is a conceptual link between causal and counterfactual claims and the contrastive counterfactual theory manages to explicate the link in a way that avoids the most dangerous pitfalls of counterfactual thinking: the theory secures the asymmetry of causal relations, does not let mere correlations to explain and does not allow backtracking counterfactual inferences to count as causal inferences. The doubts that the antagonists have presented can also be alleviated. Counterfactuals do not necessarily lead into hard-core metaphysics. The notion of "possible worlds" is not needed at all in the contrastive counterfactual theory, and even if one wanted to use the semantics of possible worlds, the semantics could be treated as an inferential tool in a way that does not involve any kind of ontological commitment. Neither does the use of counterfactuals lure one into the metaphysics of modal notions such as necessity and contingency: the questions of necessity and contingency can also be empirical questions that require counterfactual inferences and empirical evidence to be answered.

The accusations of loose semantics can also be dealt with. It is true that the sentence "If Julius Caesar had been in charge of U.N. Forces during the Korean War, he would have used (a) nuclear weapons or (b) catapults" is murky. But as already stated in chapter two the murkiness stems, in part, from the fact that based on the empirical evidence we have no clear criteria to choose between the alternatives. Secondly, the sentence is not a causal claim. When it comes to causal claims the truth makers of the claims are objective invariant dependency relations. Singular causal claims that capture objective invariant relations can be specified with the help of double contrasts '*a* [*a**] caused *b* [*b**]'. The double contrast spells out the content of the causal claims in the clearest way possible. The truth makers of causal claims are given with the help of interventionist counterfactuals:

> *a*[*a**] causes *b*[*b**] if and only if there is an intervention on the variable A such that: if the intervention changed *a* to *a**, this would bring about a change from *b* to *b**.

As it was pointed out in the fifth chapter, not all interventionist counterfactuals that fulfill the above condition are of interest to historians as causal explanations, however. On the contrary, some claims citing causes seem to be quite implausible and insensible. The obscurity of implausible counterfactuals can be alleviated with the help of the existence and mechanistic criteria that explain why some causes are considered implausible and some not. In the end, our considerations of implausible causes and counterfactuals depend on our evidence and theoretical conceptions how the world works, although the truth of interventionist counterfactuals depends entirely on whether they track down objective invariant dependency relations.

How can we then know whether our counterfactual inferences are true or not, in other words, whether they track down the objective invariant dependency relations? This question conveys the epistemic concern of the antagonists of counterfactuals. It is a good concern indeed. Although the principled claim that we cannot know the truth of any counterfactual claim is too drastic, there are cases where determining the truth of counterfactuals turns out to be guesswork. This can happen for two, often interrelated, reasons: we might not have enough evidence to decide the truth of the counterfactual or then our counterfactual inferences might become too complex when dealing with complex causal systems that we usually know only in part. These are problems that cannot be solved on the general level of philosophical theory; instead, these problems

are best dealt with case-by-case basis by trying to find new ways to handle existing evidence, to generate evidence and to make inferences. As pointed out in the fifth chapter, counterfactual inferences or "thought experiments" are not a way to generate new evidence – their quality depends entirely on the quality of evidence and the way the evidence is used.

The antagonists' attitude towards counterfactuals seems to have been partly motivated by disciplinary concerns and the desire to protect historiography's identity when faced with incursions from other disciplines. Like every discipline, historiography has its unique methods of research, but from the point of view of causal explanation historians' explanatory reasoning can be said to resemble the causal reasoning that is also employed in other fields of study. Historiography does not have its own ways of causal reasoning and explanation – also narratives are constructed and assessed with the help of counterfactual reasoning. In addition, opening up the disciplinary boundaries and borrowing theories from other disciplines can in some cases be an apt way to get new empirical evidence to support one's causal inferences.

The doubts of the antagonists are therefore answered and the argumentative gaps of the protagonists filled by the contrastive counterfactual theory. How about the methodological insights that the theory was supposed to provide? Based on the discussions of the previous chapters I have summarized the methodological guidelines provided by the contrastive counterfactual theory into a toolkit of causal reasoning in historiography:

1) When engaging in causal reasoning, specify the *explanandum* in contrastive terms: why *b* and not *b**? This clarifies your explanation-seeking question, which often helps in i) picking the right kind of causes and separating those causes from the causal background ii) turning big questions into smaller ones that are easier to answer iii) weighting causes.

2) The reliability of causal claims and the related interventionist counterfactuals depends on the strength of the evidence. Asking explicit *what if*-questions is an apt way to assess your evidence: what causal claims does the evidence support and what not?

3) If you are engaged in causal weighting specify, whether you are a) separating causes from the background factors b) doing causal weighting among causes c) analyzing the contingency of an outcome d) analyzing the sensitivity of a causal relation.

4) When engaging in counterfactual reasoning that aims to define causes of events, counterfactual inferences should follow the guidelines characterized in the fifth chapter (STEP1* – STEP4*). Be careful in trying to fulfill all the conditions of ideal interventions. The conditions set the regulative ideal for our causal inferences, but it might be that the requirements on modularity (e.g. condition 2) or the requirement on the absence of correlations (condition 3) cannot easily be fulfilled.

5) *Could*-counterfactuals that deal with possibilities should not be mixed with *would*-counterfactuals that deal with causes. *Could*-inferences are useful in the context of attributing moral responsibility and preventing unwanted outcomes, but these questions should not be used to limit the area of sensible causal claims. *Could*-counterfactuals are backward looking, whereas causal would-inferences are not. Causal inferences concentrate on what happens in the causal system after the "miracle" intervention; one does not need to worry about whether the changed caused by the intervention was actually possible or not.

It seems that applying the contrastive counterfactual theory to historiography is a methodologically fruitful endeavor that can give some concrete guidelines for causal reasoning in historiography. How does the situation look like seen from the side of the contrastive counterfactual theory and philosophy of science? Has the application of the theory into historiography been a success? Is there still room for James Woodward's doubts? To begin with the contrastive counterfactual theory easily outflanks the most discussed theory of explanation in the philosophy of historiography, namely the covering-law model. Causal explanation of human action is not a problem for this theory, either. In addition, the idea of contrastive *explanandum* seems to work in the context of historiography in a nice way and to solve some big issues like the debate of subjectivity of causal explanations. The idea that interventions work like miracles helps to distinguish *could*-counterfactuals from causal *would*-counterfactual and to see that

causal explanation is not the only context where historians use counterfactuals. In addition the theory helps to distinguish the different forms of causal weighing from each other.

Historiography also seems to point out some of the biggest weaknesses of the contrastive counterfactual theory. The implausible counterfactuals discussed by several protagonists of counterfactuals such as the counterfactual on Napoleon and Maxim machine guns reveals that not all causes are considered interesting or sensible. The theory needs to be supplemented with some extra considerations to account for this. Secondly, the idea that an intervention should directly affect only the purported *explanans* variable in a causal system of interest runs into difficulties in the context of historiography: if an ideal intervention is not possible on a system, the non-ideal intervention seems at best to provide non-ideal causal knowledge. How to characterize this non-ideal causal knowledge then? This is a question that remains to be answered. [*]

[*] I would like to thank Petri Ylikoski for introducing me to the topic of the thesis and the relevant literature. Jaakko Kuorikoski is praised for supervising my work and for his insightful and detailed comments on the script along the way. The comments I received in the seminars of professors Gabriel Sandu and Matti Sintonen helped me decide which direction to take in the early phases of the project. I am also grateful to Piiku Jakola for the last-minute language check and to Ilmari Hirvonen for his support. In addition, I will express my gratitude to the philosophy of science of community at University of Helsinki for providing me with an inspiring place to work.

References

Achinstein, Peter (1983): *The Nature of Explanation.* Oxford: Oxford University Press.

Aron, Raymond (1961): *Introduction to the Philosophy of History*, translated by George J. Irwin. London: Weidenfeld and Nicholson. The original title *Introduction à la philosophie de l'histoire* published 1938 by Librarie Galimard.

Aunesluoma, Juhana (2010): "Teoria ja historioitsijan valinnat", *Kosmopolis* 40, 2: 61–71.

Bates, Robert H.; Greif, Avner; Levi, Margaret; Rosethal, Jean-Laurent and Weingast, Barry (1998): *Analytic Narratives.* New Jersey: Princeton University Press.

Ben-Menahem, Yemima (2011): "Historical Necessity and Contingency", in in *A Companion to the Philosophy of History and Historiography*, ed. Tucker. Paperback edition (hardback edition published 2009). West Sussex: Wiley-Blackwell. 120–130.

Bogen, Jim (2004): "Analysing Causality: the opposite of counterfactual is factual", *International Studies in the Philosophy of Science* 18, 1: 3–26.

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

Bunzl, Martin (2004): "Counterfactual History: A User's Guide", *American Historical Review* 109, 3: 845–858.

Burton Adams, George (1909): "History and Philosophy of History", *American Historical Review* 14: 221–236.

Carr, David (2008): "Narrative Explanation and its Malcontents", *History and Theory* 47: 19–30.

Carr, E.H. (1964): *What is history?* Middlesex: Penguin Books.

Cederman, Lars-Erik (1996): "Rerunning History – counterfactual simulation in world politics", in *Counterfactual Thought Experiments in World Politics,* ed. Tetlock & Belkin. Princeton: Princeton University Press. 248–267.

Chisholm, Roderick M. (1946): "Contrary-to-fact-conditional", *Mind* 55, 220: 289-307.

Climo, T.A. & Howells P.G.A. (1976): "Possible Worlds in Historical Explanation", *History and Theory* 15, 1: 1–20.

Cohen, Raphael Morris (1964[1931]): *Reason and Nature: the Meaning of Scientific Method.* Glencoe: The Free Press.

Collingwood, R.G. (1994[1946]): *The idea of history*. Revised paperback edition. Oxford: Clarendon Press.

Craver, Carl (2007): *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Clarendon Press.

Croce, Benedetto (1966): " 'Necessity' in history", in *Philosophy, Poetry, History: An Anthology of Essays,* translated by Cecil Sprigge. London: Oxford University Press.

Crouzet, Francois (2001): *A History of the European Economy, 1000-2000*. Virginia: Virginia University Press.

Danto, Arthur (1965): *Analytical Philosophy of History.* Cambridge: Cambridge University Press.

Davidson, Donald (1963): "Actions, Reasons and Causes", *The Journal of Philosophy* 60, 23: 685–700.

De May, Tim & Weber, Erik (2003): "Explanation and Thought Experiments in History", *History and Theory* 42: 28–38.

Donagan, Alan (1964): "Historical Explanation: the Popper-Hempel Theory Reconsidered", *History and Theory* 4, 1: 3–26.

D'Oro Giuseppina (2011): "Historiographic Understanding", in *A Companion to the Philosophy of History and Historiography*, ed. Tucker. Paperback edition (hardback edition published 2009). West Sussex: Wiley-Blackwell. 142–151.

Dray, William (1962): "Some Causal Accounts of the American Civil War", *Daedalus* 91: 578 – 592.

— (1963): "The Historical Explanation of Action reconsidered", in *Philosophy and History,* ed. Hook. New York: New York University Press. 105–135.

— (1971): "On the Nature and Role of Narrative in Historiography", *History and Theory* 10, 2: 153–171.

— (1978): "Concepts of Causation in A.J.P Taylor's Account of the Second World War", *History and Theory* 17: 149–174.

Elster, Jon (1978): *Logic and Society: contradictions and possible worlds.* New York: John Wiley & Sons.

Fearon, James D. (1991): "Counterfactuals and Hypothesis Testing in Political Science", *World Politics* 43, 2: 169–195.

— (1996): "Causes and Counterfactuals in Social Science: exploring an analogy between cellular automata and historical processes", in *Counterfactual Thought Experiments in World Politics,* ed. Tetlock & Belkin. Princeton: Princeton University Press. 39–67.

Feinstein, Charles H. & Thomas, Mark (2002): *Making History Count: a primer in quantitive methods for historians*. Cambridge: Cambridge University Press.

Fellman, Susanna & Rahikainen, Marjatta (2012): "On Historical Writing and Evidence", in *Historical Knowledge – in quest of theory, method and evidence*, ed. Fellman & Rahikainen. Newcastle: Cambridge Scholars Publishing. 5–44.

Ferguson, Niall (2011[1997]): "Introduction", in *Virtual History: alternatives and counterfactuals*, ed. Niall Ferguson. London: Penguin Books. 1–90.

Fischer, David (1971): *Historians Fallacies: toward a logic of historical thought.* London: Routledge and Kegan Paul.

Fogel, Robert (1964): *Railroads and American economic growth*: *essays in econometric history.* Baltimore & London: The Johns Hopkins Press.

Fogel, Robert & Elton, G.R. (1983): *Which road to the past? Two views of History.* New Haven & London: Yale University Press.

Førland, Tor Egil (2004): "The ideal Explanatory Text in History: a plea for ecumenism", *History and Theory* 43, 3: 321–340.

Froeyman, Anton (2009): "Concepts of Causation in Historiography", *Historical Methods* 42, 3: 116–128.

Garfinkel, Alan (1981): *Forms of Explanation – rethinking the questions in social theory*. New Haven & London: Yale University Press.

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

Gould, J. D. (1969): "Hypothetical History", *The Economic History Review* 22, 2: 195–207.

Hanson, Victor Davis (2006): "A Stillborn West: Themistocles at Salamis 480 BC", in *Unmaking the West: "What-if?" Scenarios that rewrite world history,* ed. Tetlock, Lebow & Parker. Ann Arbor: The University of Michigan Press. 47–89.

Hart, H.L.A and Honoré, Tony (1985): *Causation in the Law*. 2nd ed. Oxford: Clarendon Press.

Hausman, Daniel M. & Woodward, James (1999): "Independence, Invariance and the Causal Markov Condition", *British Journal for the Philosophy of Science* 50: 521–583.

Hawthorn, Geoffrey (1991): *Plausible Worlds: possibility and understanding in history and the social sciences.* Cambridge: Cambridge University Press.

Hedström, Peter & Ylikoski, Petri (2010): "Causal Mechanisms in the Social Sciences", *Annual Review of Sociology* 36: 49–67.

Hempel, Carl G. (1942): "The Function of General Laws in History", *The Journal of Philosophy* 39, 2: 35–48.

— (1965): *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press.

Hempel, Carl G. & Oppenheim, Paul (1998[1948]): "Studies in the Logic of Explanation", in *Introductory Readings in the Philosophy of Science*, ed. Klemke, Hollinger and Rudge. 3rd ed. New York: Prometheus Books.

Hesslow, Germund (1983): "Explaining differences and weighting causes", *Theoria* 49: 87–111.

Horgan, Terence (1989): "Mental Quasation", *Philosophical Perspectives* 3: 47–76.

Hume, David (1748): *An Enquiry Concerning Human Understanding*, in the series *The Harvard Classics* Vol XXXXVII, Part 3. New York: P.F.Collier & Son, 1909–14; Bartleby.com 2001 URL = < http://www.bartleby.com/37/3/>

Jarausch, Konrad H. & Hardy, Kenneth A. (1991): *Quantative Methods for Historians: A guide to research, data and statistics*. Chapel Hill: The University of North Carolina Press.

Kaye, Simon T. (2010): "Challenging Certainty: the utility and history of counterfactualism", *History and Theory* 49: 38–57.

Khong, Foong Yean (1996): "Confronting Hitler and its Consequences", in *Counterfactual Thought Experiments in World Politics,* ed. Tetlock & Belkin. Princeton: Princeton University Press. 95–118.

Kosso, Peter (2011): "Philosophy of Historiography", in *A Companion to the Philosophy of History and Historiography*, ed. Tucker. Paperback edition (hardback edition published 2009). West Sussex: Wiley-Blackwell. 9–25.

Kripke, Saul (1972): *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Kuorikoski, Jaakko (2007): "Explaining with equilibria", in *Rethinking Explanantion*, ed. Persson & Ylikoski. Dordrecht: Springer. 149–162.

— (2010): *Society by Numbers: studies on model-based explanations in the social sciences*. Ph. D. thesis. *Philosophical Studies from the University of Helsinki* 31.

Landes, David (1998): *The Wealth and Poverty of Nations – why are some so rich and some so poor*. New York: W. W. Norton.

Lebow, Richard Ned (2007): "Counterfactual Thought Experiments: a necessary teaching tool", *History Teacher* 40, 2: 153–176.

— (2010): *Forbidden Fruit: Counterfactuals and International Relations*. Priceton and Oxford: Priceton University Press.

Lebow, Richard Ned & Gross Stein, Janice (1996): "Back to the Past: counterfactuals and the Cuban missile crisis", in *Counterfactual Thought Experiments in World Politics,* ed. Tetlock & Belkin. Princeton: Princeton University Press. 119–148.

Lewis, David (1973a): "Causation"*, the Journal of Philosophy* 70, 17: 556–567.

— (1973b): *Counterfactuals*. Oxford: Blackwell

— (1986a): *On the Plurality of Worlds*. Oxford: Blackwell.

— (1986b): *Philosophical Papers,* vol. 2. Oxford: Oxford University Press.
— (2000): "Causation as influence", *The Journal of Philosophy* 97, 4: 182–197.
Lipton, Peter (1990): "Contrastive explanations", in *Explanation and its Limits*, ed. Knowles. Cambridge: Cambridge University Press. 247–266.
Little, David (2010): *New Contributions to the Philosophy of History*, in the *Methodos Series,* ed. Courceau and Frank. Dordrecht: Springer.
MacDonald, Graham & MacDonald Cynthia (2011): "Explanation in Historiography", in *A Companion to the Philosophy of History and Historiography*, ed. Tucker. Paperback edition (hardback edition published 2009). West Sussex: Wiley-Blackwell. 131–141.
Mackie, J. L (1974): *The Cement of the Universe: a study of causation*. Oxford: The Clarendon Press.
Malinowski, Bronislaw (1935): *Coral Gardens and their Magic*. New York: American Book Co.
Martin, Raymond (1992): *The Past within Us: an empirical approach to philosophy of history.* Princeton: Princeton University Press.
Megill, Allan (2007): *Historical Knowledge, Historical Error: a contemporary guide to practice*. Chicago: The University of Chicago Press.
Melia, Joseph (1998): "Possible Worlds", in *Routledge Encyclopedia of Philosophy*, ed. Craig. London Routledge. URL = <http://www.rep.routledge.com/article/N088>
Mellor, Hugh (1995): *The Facts of Causation*. London: Routledge.
Menzies, Peter (2009): "Counterfactual Theories of Causation", in *The Stanford Encyclopedia of Philosophy* (Fall 2009 Edition*)*, ed. Edward N. Zalta. URL = <http://plato.stanford.edu/archives/fall2009/entries/causation-counterfactual/>.
Miklóssy, Katalin (2007): "Détente ur småstaternas perspektiv. Fallen Ungern och Rumänien", *Nordisk øst-forum,* 21, 4: 411–433.
Mitchell, Sandra (2008): "Exporting Causal Knowledge in Evolutionary and Developmental Biology", *Philosophy of Science* 75: 697–706.
Murphey, Murray (1973): *Our Knowledge of the Historical Past*. Indianapolis: Bobbs-Merrill.
Musil, Robert (1996): *The Man without Qualities,* vol I, translated by Eithne Wilkins & Ernst Kaiser. London: Vintage. Original title *Der Mann ohne Eigenschaften* (1930).
Nolan, Daniel (2011): "Why Historians (and Everyone Else) Should Care about counterfactuals", *Philosophical Studies,* published online 30 September 2011.
Northcott, Robert (2008): "Weighted Explanations in History", *Philosophy of the Social Sciences* 38, 1: 76–96.
Oakeshott, Michael (1966[1933]): *Experience and its modes.* London: Cambridge University Press.
Pascal, Blaise (1958[1669]): *Pascal's Pensées*. Paperback edition. New York: E.P Dutton & Co.
Pomeranz, Kenneth (2006): "Without Coal? Colonies? Calculus?", in *Unmaking the West: "What-if" scenarios that rewrite world history,* ed. Tetlock, Lebow & Parker. Ann Arbor: The University of Michigan Press. 241–276.
Raatikainen, Panu (2011): "Teon teoria ja historiallinen selittäminen", *Tiede ja Edistys* 36, 4: 291–300.
Raerinne, Jani: *Generalizations and Models in Ecology*: lawlikeness, invariance, stability and robustness. Ph. D. thesis. *Philosophical Studies from the University of Helsinki* 34.
Railton, Peter (1980): *Explaining Explanation: A Realist Account of Scientific Explanation*. Ph.D. Thesis. Princeton University.

Reiss, Julian (2009): “Counterfactuals, Thought Experiments, and Singular Causal Analysis in History”, *Philosophy of Science* 76: 712–723.

Roberts, Clayton (1996): *The Logic of Historical Explanation*. Pennsylvania: Penn Sate Press.

Ruben, David Hillel (1990): *Explaining explanation*. London: Routledge.

— (1994): “A Counterfactual Theory of Causal Explanation”, *Noûs* 28: 465–481.

Salmon, Wesley (1971): “Statistical Explanation”, in *Statistical Explanation and Statistical Relevance*, ed. Wesley Salmon. Pittsburg: Pittsburg University Press. 29–87.

Schaffer, Jonathan (2005): “Contrastive causation”, *Philosophical Review* 114, 3: 327–358.

Scriven, Michael (1959a): “Explanation and Prediction in Evolutionary Theory”, *Science* 130: 477–482.

— (1959b): “Truisms as the Grounds of Historical Explanations”, in *Theories of History* ed. Gardiner. New York: Free Press. 443–475.

Sintonen, Matti (1984): *The Pragmatics of Scientific Explanation*. Ph.D. Thesis. University of Helsinki. *Acta Philosophica Fennica* 37.

Steel, Daniel (2008): *Across the Boundaries: extrapolation in biology and social science.* Oxford: Oxford University Press.

Stanford, Michael (1998): *An Introduction to the Philosophy of History*. Oxford: Blackwell.

Strauss, Barry (2006): “The Resilient West – Salamis without Themistocles, Classical Greece without Salamis & the West without Classical Greece”, in *Unmaking the West: “What-if?” scenarios that rewrite world history,* ed. Tetlock, Lebow & Parker. Ann Arbor: The University of Michigan Press. 90–118.

Taylor, A.J.P (1969): *The Trouble-Makers: dissent over foreign policy*, *1729–1939*. London: Panther History.

Tetlock, Philip E. & Belkin, Aaron (eds.) (1996): *Thought Experiments in World Politics: locigal, methodological and psychological perspectives.* Princeton: Princeton University Press.

Tetlock, Philip E. & Belkin, Aaron (1996): “Counterfactual Thought Experiments in World Politics: logical, methodological and psychological perspectives”, in *Counterfactual Thought Experiments in World Politics,* ed. Tetlock & Belkin. Princeton: Princeton University Press. 3–38.

Tetlock, Philip E., Lebow, Richard Ned & Parker Geoffrey (eds.) (2006): *Unmaking the west: “What-if” scenarios that rewrite world history.* Ann Arbor: University of Michigan Press.

Tetlock, Philip E. & Parker, Geoffrey (2006): “Counterfactual Thought Experiments”, in *Unmaking the West: “What-if?” scenarios that rewrite world history,* ed. Tetlock, Lebow & Parker. Ann Arbor: The University of Michigan Press. 14–44.

Thompson, J.E.S (1966): *The Rise and Fall of Maya Civilization*. Norman: The University of Oklahoma Press.

Tucker, Aviezer (1998): “Unique Events: the underdetermination of explanation”, *Erkenntnis* 48: 59–80.

— (2004): *Our Knowledge of the Past – a philosophy of historiography*. Cambridge:Cambridge University Press.

— (2011): “Introduction”, in *A Companion to the Philosophy of History and Historiography*, ed. Tucker. Paperback edition (hardback edition published 2009). West Sussex: Wiley-Blackwell. 1–6.

Turner, Derek (2007): *Making Prehistory – historical science and the scientific realism debate*, Cambridge: Cambridge University Press.

Tuomela, Raimo (1980): “Explaining Explaining”, *Erkenntnis* 15: 211–243.

Vaidya, Anand (2011): “The Epistemology of Modality”, *The Stanford Encyclopedia of Philosophy* (Winter 2011 edition), ed. Edward N. Zalta.
URL = < http://plato.stanford.edu/archives/win2011/entries/modality-epistemology/>

van Bouwel, Jeroen & Weber, Erik (2008): “A Pragmatist Defense of Non-Relativistic Explanatory Pluralism in History and Social Science”, *History and Theory* 47: 168–182.

van Fraassen, Bas (1980): *Scientific Image.* Oxford: Oxford University Press.

Vašiček, Zneněk (2011): “Philosophy of History”, in *A Companion to the Philosophy of History and Historiography*, ed. Tucker. Paperback edition (hardback edition published 2009). West Sussex: Wiley-Blackwell. 26–43.

von Wright (1971): *Explanation and Understanding*. Itchaca and London: Cornell University Press.

Weber, Max (1949): “Objective Possibility and Adequate Causation in Historical Explanation”, in *The methodology of the Social Sciences* by Max Weber, translated and edited by Shils & Finch. Glencoe, Illinois: The Free Press.

White, Morton (1965): *Foundations of Historical Knowledge*. Westport: Greenwood Press.

Winch, Peter (1990[1958]): *The Idea of Social Science and Its Relation to Philosophy*. 2nd ed. New Jersey: Humanities Press International,

Woodward, James (1984): “A Theory of Singular Causal Explanation”, *Erkenntnis* 21: 231–262.

— (2003): *Making Things Happen*. Oxford and New York: Oxford University Press.

— (2006): “Sensitive and Insensitive Causation”, *The Philosophical Review* 115, 1: 1-50.

Ylikoski, Petri (2001): *Understanding Interests and Causal Explanation.* Ph.D. thesis. Department of Social and Moral Philosophy, University of Helsinki. URL = <https://helda.helsinki.fi/bitstream/handle/10138/21811/understa.pdf?sequence=>

— (2007): “The idea of contrastive explanandum”, in *Rethinking Explanantion*, ed. Persson & Ylikoski. Dordrecht: Springer. 27–42.

— (2011): ”Social Mechanisms and Explanatory Relevance”, in *Analytical Sociology and Social Mechanisms,* ed. Demeulenaere. Cambridge: Cambridge University Press. 154–172.

Ylikoski, Petri & Kuorikoski, Jaakko (2010): ”Dissecting Explanatory Power”, *Philosophical Studies* 148: 201–219.