LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Causal Inquiry in the Social Sciences:

the Promise of Process Tracing

Rosa Willemijn Runhardt

Submitted to the Department of Philosophy, Logic and Scientific Method for the degree of Doctor of Philosophy, on 1 April 2015.

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work.

This thesis contains material from papers that I have authored which have been published elsewhere. Firstly, the second chapter contains content from "Measurement" (Cartwright and Runhardt 2014), which was published in the volume *Philosophy of Social Science: a New Introduction*. The chapter was co-authored by Professor Nancy Cartwright and myself. I wrote all book chapter sections contained in this thesis, though my writing was supervised and occasionally amended by Professor Cartwright. I hereby declare the sections I have taken from this joint publication form an integral part of the thesis and make a relevant contribution to its main theme. Secondly, parts of chapter 3 are taken from my paper "Evidence for Causal Mechanisms in Social Science: Recommendations from Woodward's Manipulability Theory of Causation" (Runhardt forthcoming), forthcoming in *Philosophy of Science*. I am the sole author of the latter.

I declare that my thesis consists of 60,217 words.

To my parents and grandparents, if not this thesis' first then perhaps its final cause

Abstract

In this thesis I investigate causal inquiry in the social sciences, drawing on examples from various disciplines and in particular from conflict studies. In a backlash against the pervasiveness of statistical methods, in the last decade certain social scientists have focused on finding the causal mechanisms behind observed correlations. To provide evidence for such mechanisms, researchers increasingly rely on 'process tracing', a method which attempts to give evidence for causal relations by specifying the chain of events connecting a putative cause and effect of interest.

I will ask whether the causal claims process tracers make are defensible, and where they are not defensible I will ask how we can improve the method. Throughout these investigations, I show that the conclusions of process tracing (and indeed of the social sciences more generally) are constrained both by the causal structure of the social world and by social scientists' aims and values.

My central argument is this: all instances of social phenomena have causally relevant differences, which implies that any research design that requires some comparison between cases (like process tracing) is limited by how we systematize these phenomena. Moreover, such research cannot rely on stable regularities. Nevertheless, to forego causal conclusions altogether is not the right response to these limitations; by carefully outlining our epistemic assumptions we can make progress in causal inquiry.

While I use philosophical theories of causation to comment on the feasibility of a social scientific method, I also do the reverse: by investigating a popular contemporary method in the social sciences, I show to what extent our philosophical theories of causation are workable in practice. Thus, this thesis is both a methodological and a philosophical work. Every chapter discusses both a fundamental philosophical position on the social sciences and a relevant case study from the social sciences.

Acknowledgements

Everything is against the likelihood that [a text] will come from the writer's mind whole and entire. Generally material circumstances are against it. Dogs will bark; people will interrupt; money must be made; health will break down. Further, accentuating all these difficulties and making them harder to bear is the world's notorious indifference. It does not ask people to write poems and novels and histories; it does not need them. It does not care whether Flaubert finds the right word or whether Carlyle scrupulously verifies this or that fact. Naturally, it will not pay for what it does not want. And so the writer, Keats, Flaubert, Carlyle, suffers, especially in the creative years of youth, every form of distraction and discouragement. A curse, a cry of agony, rises from these books of analysis and confession. 'Mighty poets in their misery dead' – that is the burden of their song. If anything comes through in spite of all this, it is a miracle, and probably no book is born entire and uncrippled as it was conceived.

Virginia Woolf, A Room of One's Own

McAllister for first introducing me to philosophy and for subsequently encouraging me to pursue a degree in it.

At the LSE, I have had the pleasure of teaching 262 students. I know it makes you groan when I say it, but guys, believe me: the time spent teaching you was some of the best time I spent here.

Most of all, though, I want to thank my husband Yorick van Boheemen-Runhardt for being there for me from my first week in university, now nearly ten years ago. He neither needs philosophy nor cares for it, but he has kept the dogs from barking, the people from interrupting, and my health from breaking down.

Contents

Chapter 5 – Epistemic Homogeneity: a Taxonomy

Chapter 6 – Thinking about Social Mechanisms

Chapter 7 – Conclusion

Chapter 1

Introduction

1. Setting the scene

What causes civil wars? Why did insurgents start using terrorist tactics in the Second Chechen War? Why did some Salvadoran peasants join the FMLN resistance fighters, while others refused? Are ethnic wars really caused by different factors than resource conflicts? What made the French government decide to back down after the British threatened to use force over land control in Egypt in 1898? These questions are all about *causes* in the social world.

In this thesis, I ask what evidence we need to answer such causal questions. In particular, I investigate a recent method that attempts to give evidence for a causal relation by specifying the chain of events connecting a putative cause and its effect of interest, 'process tracing', and show whether it succeeds or not. I investigate whether the causal claims process tracers make about the cases they study are defensible; where they are not defensible I ask how we can improve process tracing. Throughout these investigations, I will show that the conclusions of process tracing (and indeed of the social sciences more generally) depend not only on the causal structure of the social world, but also on social scientists' aims and values.

My central argument will be this: all instances of social phenomena have causally relevant differences, which implies that any research design that requires some comparison between cases (like process tracing) is limited by how we systematize the phenomenon. Moreover, such research cannot rely on stable regularities. Nevertheless, to forego causal conclusions altogether is not the right response to these limitations; by carefully outlining our epistemic assumptions we can make progress in causal inquiry.

While I use philosophical theories of causation to comment on the feasibility of a social scientific method, I also do the reverse: by investigating a popular contemporary method in the social sciences, I will show to what extent our

philosophical theories of causation are workable in practice. Thus, this thesis is both a methodological and a philosophical work. Every chapter discusses both a fundamental philosophical aspect of the social sciences and a relevant case study in the social sciences.

There are several aspects of the literature that I will evaluate in this introduction to set the scene for later chapters. In the first part of the introduction, I will present a taxonomy of different philosophical theories about causal inquiry and link them to several methodological approaches. I also outline several of the key problems for causal inquiry in the social sciences. Having introduced these parts of the philosophical literature, I turn to the methodological literature. I give a brief history of methods for causal inquiry, emphasizing the distinction between 'correlation accounts' and 'process accounts'. I show how process tracing was developed, highlighting the problems that process tracing was meant to solve, and finish with an overview of the main texts on the topic.

In the second part of the introduction, I discuss two criteria for good social science research methods, which will guide my analysis of process tracing in this thesis: transparency (i.e. whether the method gives evidence for intermediate factors), and the comparability of cases under study.

In the third part of the introduction, I outline my main arguments in the thesis regarding the usefulness as well as the fallibility of process tracing. I present my key strategies and assumptions, as well as potential limitations of my analysis.

Finally, I give brief abstracts for the five chapters in this thesis and show how the chapters link to the themes discussed in parts 2 and 3 of the introduction.

2. A taxonomy of causal inquiry

2.1 Philosophical positions on causal inquiry

A central question underlying causal inquiry is what it means for C to be a cause of E. There exists a variety of definitions of causation in the literature: for some, C and

E are causally related if *C* raises the probability of *E*; for others, *C* and *E* are causally related if they are connected by a chain of intermediate events; yet others define the causal relation in terms of counterfactuals, manipulation, or capacities. Such metaphysical theories of causation in turn imply distinct epistemological and methodological theses: as we will see, whether one prefers large-N, statistical work to a small-N, case-based approach depends in part on what one considers causation to be.

Many authors have presented their own taxonomy of the different theories of causation in the social sciences (Brady 2008, Bunge 1997, Cartwright 2002, Dupré and Woodward 2013, Gerring 2005, Humphreys 1986, Kincaid 2009, Little 1991, Reiss 2009a, Tilly 2001). Generally, such taxonomies are written for the social sciences only (rather than encompassing causal relations more generally), since it is believed that causal inquiry in the social world has its own unique challenges and limitations. In their taxonomies, philosophers and methodologists generally distinguish between the following theses about causation:

Counterfactual accounts. Relying on counterfactuals means using 'what if' scenarios to identify the causes of certain events of interest. We may ask, for instance, what would have happened if Archduke Franz Ferdinand had not been assassinated in order to investigate whether his assassination was one of the causes of the First World War. Because we cannot repeat history to see what *actually* would have happened if Franz Ferdinand had lived, we must rely on 'sufficiently similar' cases. So, arguably, any evidence for a counterfactual relies on some similarity comparison. Counterfactual accounts form the basis for a variety of social science methods, including historical thought experiments (cf. Reiss 2009b).

Interventionist and manipulation accounts. These accounts, which are sometimes seen as a special case of counterfactual accounts (cf. Psillos 2004), claim that *C* is a cause of *E* if there exists some intervention *I* by means of which we can manipulate *C* which in turn will affect *E* (cf. Woodward 2003). For instance, if classroom size negatively affects students' exam results, then decreasing class size will increase students' exam results. Interventionist or manipulation accounts are

often related to experimental methods (cf. Brady 2008), even though experiments are difficult to devise for the social sciences (cf. Humphreys 1986). Others argue that devising hypothetical experiments can in some cases be sufficient to give evidence for a causal claim (cf. Woodward forthcoming).

Capacities accounts. In these accounts (cf. Cartwright 1989), a factor X (e.g. youth unemployment) has the capacity to Ψ (e.g. raise the probability of civil war) if and only if X has the ability to Ψ even when there are factors that prevent X from actually producing Ψ (e.g. when there are hindering factors that prevent civil war from being more probable). One condition for X to have the capacity to Ψ is that X's ability to Ψ is stable across a range of background conditions. There is some discussion as to whether there exist capacities in the social world, since many social scientific methods do not give evidence for capacities; after all, the argument goes, if the conclusions of our research are tangled up with test conditions, then we cannot conclude anything about how X would work in isolation (cf. Reiss 2008b).

Regularity accounts. According to regularity accounts, a cause is always followed by its effect (cf. Hume 1748). Thus, a factor is a cause of a phenomenon if and only if it is "a member of a group such that that group of factors is always associated with the phenomenon of interest and no subgroup is always associated with the phenomenon" (Reiss 2009a, 23). Regularity accounts are associated with such techniques as qualitative comparative analysis or QCA (cf. Ragin 2008). In QCA, one partitions a small sample of cases into a typology of subgroups according to those cases' properties. The researcher then tries to interpret any clustering together of properties across cases.

Probabilistic accounts. In probabilistic accounts of causation, C is a cause of E if and only if, holding any background factors B fixed, C changes the probability of E. A general causal claim like 'smoking causes cancer' is seen as a conditional probability statement $P(E|C,B) > P(E|B)$. Probabilistic accounts are associated with statistical techniques like linear regression analysis, and with philosophical frameworks like causal modelling (Pearl 2000, Spirtes, Glymour, and Scheines 1993). Probabilistic accounts should not be interpreted lightly; without sophisticated

accounts of confounding factors and common causes, they can seem to indicate that correlation implies causation. However, correlation is neither necessary nor sufficient for causation (cf. Reiss 2009a).

Mechanistic accounts. A mechanistic account of causation in the social sciences implies that C is a cause of E if and only if there is a mechanism connecting C and E. Such a mechanism produces an observable chain of events leading from C to E. Some authors believe that a mechanism connecting a cause and effect should be based on some kind of law or regularity. Daniel Little, for instance, defines that C is a cause of E if and only if "there is a series of events C_i leading from C to E, and the transition from each C_i to C_{i+1} is governed by one or more laws L" (Little 1991, 14). Others argue that mechanisms do not have to be lawlike (Bogen 2005). Mechanistic accounts have not only become more prominent in the social sciences, they are also widely discussed in philosophy of biology (cf. Machamer, Darden, and Craver 2000).

2.2 Complications particular to the social sciences

The accounts outlined above all give an interpretation of what it means for some C to cause E. We ought to separate this question from another important issue in philosophy of causation, i.e. what kind of evidence can *corroborate* causal relations. Here, I want to outline some of the main challenges that researchers in the social sciences face when trying to give evidence for causal claims.

The first issue for the social sciences, as opposed to other sciences, is that we are often dealing with aggregate, social, non-individual entities (e.g. institutions, like courts, hospitals, and universities). It is up for debate how the properties of these entities are related to the properties of individual agents. Methodological individualists have traditionally argued that "we shall not have arrived at rock-bottom explanations of (…) large-scale phenomena until we have deduced an account of them from statements about the dispositions, beliefs, resources, and interrelations of individuals" (Watkins 1957, 442). On the other hand, sociological

holists have argued that "social systems constitute 'wholes' at least in the sense that some of their large-scale behaviour is governed by macro laws which are essentially sociological in the sense that they are *sui generis* and not to be explained as mere regularities or tendencies resulting from the behaviour of interacting individuals" (Watkins 1957, 442). We may ask what this means for causal inquiry; Richard Miller, for instance, has accepted that "[e]very social phenomenon is caused by the acts of individuals" (Miller 1978, 470), but argued this does not imply methodological individualism.

The second issue for the social sciences is that social phenomena are often the result of a complex set of interacting causes working at the same time, i.e. we are often dealing with multiple causation. Additionally, we can often describe causes at work in the social sciences as INUS conditions, i.e. insufficient but necessary parts of an unnecessary but sufficient condition (cf. Mackie 1974). Such causes need to work together with other factors to lead to the effect, but there are other combinations of factors which can lead to the effect too.

Third, and relatedly, even if we manage to isolate one particular cause C to judge what its relation to effect E is, the other causes (background conditions) that enable or hinder the causal relation between C and E are often unstable (cf. Cartwright and Efstathiou 2011). Any phenomenal regularity in the social world is likely to change as the structure of society changes; the worry is that this may imply that social phenomena are "in principle unpredictable" (Little 1993, 186). Moreover, the interventions we may use to test a particular relation $C \rightarrow E$ can, in the social sciences, be structure altering, meaning that they do not only influence the cause and effect but also some causally relevant background conditions. How this impacts on research methods is a matter of debate (cf. Cartwright 2001, Woodward 2008).

The above issues with social causation have led to several particular demands on good research design for finding causes in the social sciences. I will discuss these in part 3 of this introduction. First, however, let me turn to the history of the method of causal inquiry this thesis is centred on: process tracing.

2.3 Political science perspectives on causal inquiry: from correlations to mechanisms

The process tracing method was initially proposed by political scientists who rejected what they called a more traditional 'neopositivist' or 'correlational' view of causation (cf. Checkel 2008, 125): the idea that "a factor is the cause of an outcome (…) [if] the factor is systematically associated – constantly conjuncted, or covaried – with it" (Jackson 2011, 68), and that "a causal connection shows itself in systematic cross-case correlations between specific factors" (Jackson 2011, 41).

One of the key methodological texts that subscribes to this neopositivist view is Gary King, Robert Keohane, and Sidney Verba's *Designing Social Inquiry* (King, Keohane, and Verba 1994). King, Keohane, and Verba's key concept of causation, based on earlier work by Paul Holland (1986), is the 'causal effect', i.e. the value of an effect variable if the cause is present relative to its value if the cause is not present (King, Keohane, and Verba 1994, 85). King, Keohane, and Verba suggest that this view of causal inquiry should underlie both what they call 'quantitative' and 'qualitative' research methods; this means that they treat all causal inquiry as "essentially statistical" (McKeown 1999, 169).

Designing Social Inquiry has since come under scrutiny for its statistically inspired arguments about case study methods. For instance, the authors' claim that case study research suffers from a degrees of freedom problem (where there are more variables being tested than observations to test them) is said to ignore non-statistical approaches to causal inquiry, e.g. the possibility of drawing on general principles or historical facts to establish a causal relation (McKeown 1999, 169-170). Moreover, instead of arguing, as King, Keohane, and Verba had done, that one needs correlations to give evidence for causation, case study researchers emphasize the use of causal mechanisms. The distinction between 'correlational causal analysis' and 'the analysis of causal mechanisms and processes' thereby became a key issue in political science methodology (cf. Kincaid 2009, 739-740).

We can see this shift to mechanisms in the methodological literature most clearly from the early 2000s onwards. In 2001, Charles Tilly recognized the following trend in political science:

> Despite more than a century of strenuous effort, political scientists have [not] securely identified [any general uniformities in the propensities of human actors]. But they have recurrently identified widely operating causal mechanisms and processes. Rather than continuing to search for propensity-governing covering laws, it would therefore make sense to switch whole-heartedly toward specification of mechanisms and processes. (Tilly 2001, 25)

James Mahoney, in the same year, also recognized a shift in techniques, highlighting claims by earlier authors like Hedström and Swedberg (1998) that "correlational analysis is by itself an inadequate mode of causal assessment" (Mahoney 2001, 575). According to Mahoney, one of Hedström and Swedberg's main reasons for this claim is that "correlational analysis fails to identify the mechanisms that produce observed associations and that explain why empirical associations exist in the first place" (Mahoney 2001, 576). Similarly, Peter Hall discusses the common view that "the key characteristic of a theory is that it does not simply identify an empirical regularity, but adduces reasons as to why this empirical regularity should exist, and setting out those reasons usually entails outlining causal mechanisms associated with the phenomenon at hand" (Hall 2013, 21). In other words, these authors claim that one ought not to focus on tenuous causal claims between distant variables, but rather on what connects these distant variables.

King, Keohane, and Verba's critics, who sometimes call themselves 'modern constructivists' or 'scientific realists' (not to be mistaken with the view of realism in general philosophy of science), are committed to projects which produce a list of causal mechanisms for a particular case study, and then try to bring in theory to generalize to a wider set of cases. Such generalization amongst others involves specifying the 'scope conditions' of a case study, i.e. showing that the mechanisms work outside our immediate area of study. These scope conditions are often

supported using cross-case analysis, which involves comparing several case studies. These case studies can be cross-national (i.e. across different countries) or subnational (i.e. across different regions).

The modernist constructivists also focus on another sense in which their work is 'general': they claim that they typically attempt 'middle range' research. For instance, the modernist constructivists suggest they would not simply look at a specific civil war in a specific country at a specific time, but also not at civil wars in general. 'Middle range' here means investigating conflict at some level of abstraction between these two, like 'civil wars in Europe in the late twentieth century'. Arguably, how to delineate the scope conditions of a theory (to how many cases a theory applies) is a different question from the matter of what level of abstraction is appropriate for a theory. Though I investigate the former in detail in this thesis, I will discuss the latter only indirectly.

2.3.1 The potential benefits of mechanistic reasoning

There are several (philosophical) arguments in the literature for why focussing on mechanisms may be preferable to focussing on correlations (cf. Johnson 2006, Little 1991, Reiss 2007, 2009a, Steel 2004). Firstly, a probabilistic claim that $P(E|C,B) > P(E|B)$ for some background context B (which we need to prove that C and E are causally related, according to the proponents of probabilistic causation) is very difficult to evaluate in an area like the social sciences. After all, in areas of complex causation, specifying B is going to be difficult.

Secondly, large-N analyses have a tendency to average out differences between individual cases; however, if there are always going to be causally relevant differences between the different cases of some phenomenon (e.g. if some civil wars are much more strongly influenced by the country's level of youth unemployment than other civil wars are), then looking at the averages is not useful.

Thirdly, there is no prima facie reason to believe that even the most stable phenomenal regularities in the social world (think e.g. of the democratic peace

theory, the theory that no democracies go to war with each other) correspond to some deeper causal regularity. We may therefore accept Daniel Little's claim that "the central explanatory task for social scientists is to uncover causal mechanisms, not to formulate explanatory regularities that permit the deduction of observed phenomena" (Little 1993, 185). Note that the latter does not mean we reject *all* regularities in social science; indeed, as Little continues in the same passage, "[t]here are regularities that correspond to causal mechanisms, to be sure; but these may not be discernible (because of the difficulty of isolating causal factors), and they are unlikely to take the form of strong high-level regularities across social contexts" (Little 1993, 185). I will discuss the relation between regularities and mechanisms in most detail in chapter 6.

2.4 The development of process tracing

One of the main methods of modernist constructivists is process tracing. Process tracing involves measuring the observable consequences of causal mechanisms. We can see its development start in the early 2000s. In 2004, Henry Brady and David Collier published an edited volume of responses to the King, Keohane, and Verba framework, *Rethinking Social Inquiry* (Brady and Collier 2004). While King, Keohane, and Verba, as we have seen, emphasize estimating the causal effect, Brady and Collier's volume tries to show what lies behind correlations. They focus, in light of the modernist constructivist tradition I outlined above, on causal mechanisms. This also means that while King, Keohane, and Verba emphasize cross-case comparisons, Brady and Collier present tools for within-case analyses. Brady and Collier's work was received by some as a 'corrective', elucidating some of the aspects of causal inquiry that King, Keohane, and Verba did not cover in sufficient detail, but not presenting an alternative to their methods (cf. Gerring 2008a).

In 2005, Alexander George and Andrew Bennett continued the mechanist trend by publishing their own more extensive response to the neopositivist views in King, Keohane, and Verba, *Case Studies and Theory Development in the Social Sciences*

(George and Bennett 2005). This work *did* present an alternative to King, Keohane, and Verba, by analyzing case study methods in detail. It was also one of the first, and now most commonly cited (cf. Bennett and Elman 2006, 2007, Collier 2011a, b, Hall 2013, Waldner 2012), systematic presentations of process tracing, the result of a longer working relationship between George and Bennett (see for instance Bennett and George (1997), and earlier independent work by George (George 1979)).[1]

In the book, George and Bennett argue that "causal explanations must include two things: correlational or probabilistic statements associating purported causes with observed effects, *and* logically coherent and consistent assertions on the underlying causal mechanisms through which purported causes affect outcomes" (George and Bennett 2005, 39). Because, according to the authors, statistical methods are "not well suited to testing causal mechanisms in the context of particular cases" (George and Bennett 2005, 44), they must be complemented with process tracing.

Though George and Bennett brought process tracing into the mainstream, it took another decade before an entire book was published on the method. Alexander George having passed away in 2006, Andrew Bennett published his next work in cooperation with political scientist Jeffrey Checkel. In their co-authored volume *Process Tracing: from Philosophical Roots to Best Practices*, Bennett and Checkel presented a systematic analysis of the philosophical foundations of process tracing, "a philosophical base that is ontologically consistent with mechanism-based understandings of social reality" (Bennett and Checkel 2015, 21). The authors argue for process tracing as the best tool to find evidence of causal mechanisms, and they establish evaluative standards to judge what makes any particular study an instance of good process tracing.

Bennett and Checkel, following George and Bennett's earlier systematization, define process tracing as "the analysis of evidence on processes, sequences, and conjunctures of events within a case for the purposes of either developing or testing

---

[1] In the second edition of Brady and Collier's volume (Brady and Collier 2010), a chapter by Andrew Bennett was added, outlining the process-tracing technique.

hypotheses about causal mechanisms that might causally explain the case" (Bennett and Checkel 2015, 7). Herein, a case is defined as "an instance of a class of events" (George and Bennett 2005, 18). In other words, process tracing is used to find out how, within the temporal, spatial, or topical boundaries of a particular case, a particular outcome we are interested in came about.

Process tracing, according to Bennett and Checkel, can be both 'bottom up' (i.e. develop hypotheses) and 'top down' (i.e. test "the observable implications of hypothesized causal mechanisms within a case" (Bennett and Checkel 2015, 7)). Bottom-up process tracing, Bennett and Checkel point out, is usually preceded by a period of immersion in a case to collect extensive information and come up with potential explanations. It is only once sufficient information has been gathered that a researcher can move on to top-down process tracing, which tests process-tracing hypotheses and theories. One of the important stages in top-down process tracing is "to develop case-specific observable implications of the theories in question" (Bennett and Checkel 2015, 18).

3. Central themes for analysing methods of causal inquiry

So far I have presented a selective history of process tracing, as well as some of the motivations for using a mechanism-based method of social inquiry rather than, or in addition to, a correlation-based method. I will now set out the central themes of this thesis. My central aim is to evaluate process tracing as a method of causal inquiry in a series of interrelated chapters. In this section, I will present the two criteria on the basis of which I will evaluate process tracing. These criteria will prove to be relevant not only for process tracing but also for other social science research methods; occasionally therefore my conclusions in subsequent chapters will be relevant to other methods as well.

The first criterion I will discuss is *transparency*. This criterion demands that the research method gives an insight into the process connecting cause and effect. The second criterion is *comparability*. This criterion demands that the samples under

study in the research design contain similar enough individuals or cases to draw conclusions.

My terminology derives from John Gerring, who argues for these criteria and several others (Gerring 2005). These two criteria are not meant to be comprehensive; rather, they provide me with a structured way of moving the debate about process tracing forward. I have chosen these two criteria because they (as I will show below) have been highlighted by a variety of authors, both in philosophy of science and in methodology. I consider them not only important norms, but also descriptive of the aims of social science researchers.

3.1 Transparency

The first criterion for good research design that I will use to structure my thesis is that of *transparency.* By this, Gerring simply means to ask the following: "Does the research design offer evidence about the process (i.e. the intermediate factors) by which X affects Y?" (Gerring 2005, 183). At first glance, process tracing fits the bill. I will show that what is especially important in analysing process tracing's transparency is to detail exactly what we mean by a process connecting a cause C and effect E. There exist several alternative interpretations in the literature.

The first interpretation of a process is exactly how Gerring defines it, i.e. in terms of intermediate factors. Yet, to consider a process as a set of intermediate factors opens one up to criticism of infinite regress. After all, after having found the set of intermediate factors $Z_1, Z_2, …, Z_n$ between C and E, we then have to provide a set of intermediate factors between each Z_i and Z_j. This is also King, Keohane, and Verba's criticism of causal mechanisms:

> If we posit that an explanatory variable causes a dependent variable, a 'causal mechanisms' approach would require us to identify a list of causal links between the two variables. This definition would also require us to identify a series of causal linkages, to define causality for each pair of consecutive variables in the sequence, and to identify the linkages between

any two of these variables and the connections between each pair of variables. This approach quickly leads to infinite regress, and at no time does it alone give a precise definition of causality for any one cause and one effect. (King, Keohane, and Verba 1994, 86)

Yet, there is more to a process connecting a cause C and effect E than just the chain of intermediate factors. Other authors have argued for an interpretation of this chain of events that depends on one of the other fundamental philosophical theories of causation, like counterfactual reasoning (Psillos 2004) or regularities of some kind (Machamer, Darden, and Craver 2000). Such alternative interpretations arguably avoid the problem of infinite regress: if we can find evidence of a connection between Z_i and Z_j that does not rely on another chain of intermediate factors, but on e.g. a counterfactual or a regularity, we do not need to look for further chains of events connecting Z_i and Z_j.

However, although philosophers of science have been considering alternatives to the intermediate-factor interpretation of processes, the methodological literature is relatively silent on the issue. There is some attempt at showing how causal mechanisms and processes relate (Ylikoski 2011), but since there is also a wide variety in how the term 'mechanism' is defined and used (cf. Mahoney 2001), work still remains to be done to show what, if anything, adopting a more fundamental theory of causation can contribute to process methodologies. In chapter 3, I will investigate what adopting James Woodward's interventionist theory of causation can contribute to process methodologies. In chapter 6, I investigate what Machamer, Darden, and Craver's theory of causal mechanisms in biology (Machamer, Darden, and Craver 2000) can contribute.

3.2 Comparability

The second criterion for good research design that will guide my thesis is comparability. With comparability, Gerring attempts to capture the worry that in order to draw general conclusions from case study research, the case must be

subsumable under a wider category, i.e. it must be similar to other cases in some respects. For instance, if we wish to make a general claim about a sample which includes both the Iraqi and the Syrian civil wars, that means that the Iraqi and Syrian civil war must be sufficiently similar.

The comparability criterion, Gerring argues, comprises two separate criteria: *descriptive* comparability and *causal* comparability. Descriptive comparability means that the cause and effect in the cases in the sample must refer to the same things; it is unhelpful if the cause in the first case is slightly different from the cause in the second. To come back to our example, to give evidence for the claim that youth unemployment is one of the causes of civil war, we must not only mean the same by 'civil war' in the Syrian and in the Iraqi case, but what we mean by 'youth unemployment' in Syria must also be similar to what we mean by 'youth unemployment' in Iraq. Causal comparability on the other hand is meant to exclude situations where there are some idiosyncrasies about the way the cause and effect are related for one or more of the cases. For instance, if youth unemployment was a more influential factor in Syria than it was in Iraq, then we would break the causal comparability criterion.

It is worth noting that in Gerring's framework, having some way of controlling for any descriptive or causal dissimilarities between cases in the sample is also important. Here, I will highlight descriptive and causal comparability to show what issues the social sciences face; I will deal with the solution, including controlling for dissimilarities, in the relevant chapters in the thesis itself.

3.2.1 Descriptive comparability

John Gerring describes descriptive comparability as the comparability of the *X*s (the causal factors) and *Y* (the effect of interest) "such that '*X*' and '*Y*' mean roughly the same thing across cases" (Gerring 2005, 184). In other words, "each observation [in the study] must be score-able on some scale, and the attribute must mean (roughly or precisely) the same thing across the contexts in which it is being compared"

(Gerring and Thomas 2011, 8). Without descriptive comparability, one cannot draw general conclusions about a set of cases.

As Gerring argues, large-N work presumes that its cases are highly comparable; it requires a single, uniform measure of the Xs and Y. If the cases which large-N studies generalize over are not similar, one averages out over differences. Similarly, mid-range theories require comparability as well. (On the other hand, a within-case study does not have to rely on descriptive comparability, since it only discusses the Xs and Y for one case; though that is not to say it does not require a careful systematization of what variables are under discussion.)

How, then, do we find out whether a research method satisfies descriptive comparability? Finding descriptive comparability is particularly difficult to accomplish in the social sciences, and this highlights one of the complications with causation that I pointed out in the above. In the social sciences, we are dealing with aggregate, social, non-individual entities (cf. Kincaid 2009). Think, for instance, of the concept 'democracy'. Not all democracies are alike, yet there are researchers who aim to make causal claims about all democracies (e.g. the democratic peace theory).

One of the key issues in this context is how to systematize fuzzy social science concepts. This issue is widely discussed in the literature on philosophy of measurement (cf. Cartwright and Runhardt 2014). Think, for instance, of 'poverty'. Though we have an intuitive understanding of what a concept like poverty means, in order to make general claims about poverty we need to characterize the term (e.g. with a set of necessary and sufficient conditions, telling us what criteria someone must meet to be considered poor), represent it (e.g. on a scale, or as a binary variable), and come up with procedures to then decide whether a particular case is an instance of the concept ('is this particular person poor?'). Philosophy of measurement so far has not given sufficient attention to how the causal structure of phenomena constrains how one ought to systematize such phenomena; this will provide the starting point of my analysis in chapter 2.

3.2.2 Causal comparability

The second kind of comparability that Gerring highlights is causal comparability, i.e. "choosing cases that are similar to each other in whatever ways might affect the Y or the posited $X \rightarrow Y$ relationship" (Gerring 2005, 185). We can find this demand in other authors, as well, under the heading of 'unit homogeneity'. To demand unit homogeneity means to ask that "all units with the same value of the explanatory variables have the same expected value of the dependent variable" (King, Keohane, and Verba 1994, 91). It is worth noting that causal homogeneity presumes descriptive homogeneity; without being able to compare cause and effect variables in two cases, we cannot compare their causal relationship.

For the social sciences, and indeed more generally, causal comparability is difficult to achieve because one almost never has a complete causal picture of the phenomena under study (cf. Steel 2008), especially given the issues of multiple causation and INUS factors highlighted in section 2.2. Without knowing what kinds of factors may interfere with Y other than X, it is difficult to see whether some value of X will have the same effect on the value of Y for two cases. For that reason, some authors have argued that comparability or homogeneity is too strong a demand on social science methods (cf. Hausman 2010). Though we may take the lack of comparability as a reason to stick with single case research and avoid generalization altogether, this solution is unsatisfactory for some purposes; causal comparability is required for any kind of generalization, and generalization in turn is important for both theorizing and policy purposes. In chapters 4 and 5, I provide a new analysis of causal comparability that takes into account our often limited knowledge of the causal structure of social phenomena.

3.2.3 Representativeness and external validity

Note that in Gerring's list of requirements for good research design, he also discusses a criterion that I will discuss in this thesis as an extension of the demand for comparability: *representativeness*. With representativeness we are talking not of

the comparability of cases within our sample, but about comparability between sample cases and some target cases. Representativeness becomes an issue whenever we wish to conclude something about a range of cases wider than those we have studied. Whether one may generalize from cases under study to a set of target cases depends on whether the causal factors and relationship in the sample are comparable to those of the cases we wish to extrapolate to.

In the philosophy of science literature, issues of representativeness are sometimes discussed under the header of 'external validity' (Cartwright 2007, Cartwright and Efstathiou 2011). The key argument in the external validity debate is that following a good method of causal inquiry alone cannot guarantee us that the results we have found are representative for a wider population. For that, we require additional information about the target cases. Yet, problematically, this information can only be collected by studying the target cases in as much detail as we have studied the test cases. In Gerring's words, "no methodological procedure will overcome this basic assumption [of representativeness], which must be dealt with in light of what we know about particular phenomena and particular causal relationships." (Gerring 2005, 187)

Epistemic strategies for dealing with external validity issues include Daniel Steel's 'comparative process tracing'. This technique (which we should distinguish from the process tracing method under discussion in my thesis) was developed by Steel in reference to external validity issues in biology in the first instance, though Steel also tries to apply the technique to the social sciences. The key idea in this technique is to compare stages of the mechanism in the test case to those in the target case, focusing only on the parts of the mechanism that may be significantly different. Researchers should judge which parts these are "based on inductive inferences concerning known similarities and differences in related mechanisms in a class of [cases], and the impact that those differences make" (Steel 2008, 89). This is easier in a field like biology than it is in the social sciences, as indeed Steel himself admits; in the social sciences, there are most likely always going to be causally relevant differences between a test case and a target case.

Using my analysis of both descriptive and causal comparability, I will comment on the issues with representativeness in chapters 4 and 5.

4. Argument

In this thesis, I will show whether process tracing can meet the above two criteria (transparency and comparability), and thereby also comment on how they are related. I spend time going into detail on both criteria, showing what they mean for social scientific practice in general, and how they relate to the debate regarding process tracing. Although there is a main line of argument running through the thesis, each chapter can be considered independently as well.

Before I present a summary of what I will do in the rest of this thesis, I wish to outline several assumptions I have made. Firstly, every chapter of this thesis contains a case study to illustrate my main argument. These case studies are all from the field of international relations or political science, and many are taken particularly from conflict studies. This makes the thesis more focused, since it allows me to draw comparisons between different approaches in one particular field. Moreover, almost all of the methodologists whose ideas I discuss here are themselves working in international relations or political science; Andrew Bennett and Jeffrey Checkel, for instance, both work as professors in international relations, Alexander George was professor of political science, and Henry Brady, David Collier, and James Mahoney are all associated with political science departments.

This narrowness of focus opens my work up to the criticism that if what I argue is true for the particular area of the social sciences I have been looking at, this does not automatically imply my conclusions are true for the social sciences more generally. This thesis may be considered, by those critics, a work in philosophy of international relations or political science. I strongly believe that many of the underlying issues I discuss here (particularly the issues of transparency and comparability) are issues not just for international relations but for the social sciences more generally. The fundamental philosophical ideas I will present here

are based on philosophy of the social sciences more generally, not just on philosophy of international relations or political science. Nevertheless, I believe this criticism indicates that further work can be done to show how my analysis applies in other areas of the social sciences, such as sociology. I will come back to this issue in my conclusion.

A second decision I have made in this thesis is to focus on social science practice, rather than more fundamental ontological questions. I ask myself how causation works in the social world only for as far as this has consequences for how one can find evidence about causation. (As such, for instance, I do not concern myself with the question of methodological individualism versus holism, which I briefly outlined in section 2.2.) There is one aspect of this decision that I want to discuss here, viz. my assumption of 'causal pluralism', which has come under criticism in the recent literature (Reiss 2009a).

In my thesis, I will be considering different approaches (mechanisms, but also probabilistic reasoning, manipulability, and counterfactuals) to come to a conclusion about what works best for causal inquiry in the social sciences. I will show, for instance, that evidence of counterfactuals can support (mechanism-based) process tracing, and I will comment on how evidence of mechanisms can help us improve probabilistic causal claims. This 'mixed methods' approach, arguably, assumes evidential pluralism, i.e. the view that there exist different, compatible kinds of evidence for causal relations. It fits in with the move Julian Reiss describes as

> to loosen the relationship between causation and what one might call the 'manifestations' of causation such as counterfactual dependence, correlation, stability under intervention, and so on. The manifestations of causation, according to this response, are not regarded as *defining* causation or as expressing *characteristics universally associated with* causal relationships but rather as providing *evidence* or *test conditions* for the existence of causal relationships (Reiss 2009a, 27, emphasis in original).

Although I believe that this pragmatic decision to draw on different sources of evidence in the social sciences helps move process tracing forward, and I hope this will become clear in the chapters that follow, my work is nevertheless susceptible to the following worry. Julian Reiss argues for the strong claim that "having evidence in favour of a causal claim of one type [e.g. mechanistic causation] does not (…) entitle the bearer of the evidence to the belief in another type of causal claim [e.g. probabilistic causation], even tentatively" (Reiss 2009a, 34). Different kinds of evidence, in Reiss's view, "tend to support different types of causal claim (…). [T]he different kinds of causal claim are sometimes true of the same system, but whether that is so is an empirical question that has to be addressed, and answered supported by evidence, in its own right" (Reiss 2009a, 37). Thus, in order for my reliance on different kinds of evidence to be successful, I ought to show that for the systems I investigate, the different kinds of causal claim are compatible. In what follows, I will do so by carefully examining the agreements and differences between different kinds of causal claims for each case study. I will leave the question of whether I have been successful to my conclusion.

4.1 Chapters in this thesis

This thesis consists of a set of interconnected chapters that each treat aspects of the two requirements for good research methods, and which are based on the assumptions and limitations just mentioned. As I anticipated above, my central argument will be this: all instances of social phenomena have causally relevant differences, which means that any research design that requires some comparison between cases (like process tracing) is limited by how we systematize the phenomenon, and cannot rely on stable regularities. Nevertheless, to forego causal inquiry altogether is too strong a reaction; by employing epistemic strategies and carefully outlining assumptions we can make progress in uncovering causal relations.

Chapter 2 – Measuring Civil War: the Causal Constraints on Systematizing Social Science Phenomena

I begin, in chapter 2, by laying a foundation for my discussions of causation by considering what descriptive comparability entails in the social sciences. I discuss a common position in philosophy of social science, 'pluralism', that maintains that certain social science terms have a variety of meanings across different contexts. Examples of such terms include 'well-being', 'disability', and 'democracy'. Pluralists argue the way the terms are defined depends on moral, prudential, political, and social values. In the chapter, I set out one constraint on this position: I outline a case in which it is not just values, but also causal inquiry that constrains our definition of the term. In a detailed case study, viz. the development of the large-N Collier-Hoeffler theory of the causes of civil war, I identify a number of causal assumptions about the nature of civil war that have informed the researchers' definition of the term 'civil war'. I conclude, more generally, that although social science phenomena can and do get characterized with different opposing definitions, nevertheless which definitions we allow depends on the causal structure of the phenomena. A direct consequence of this argument is that one can make epistemic progress regarding phenomena that are defined in multiple ways.

Chapter 3 – Evidence for Causal Mechanisms in the Social Sciences: Recommendations from the Manipulability Theory of Causation

In the next chapter, chapter 3, I first introduce process tracing, and consider whether it meets the transparency criterion outlined above. Process tracing claims to investigate the chain of events connecting a putative cause and effect, as the transparency criterion requires. However, as I will show, simply listing a chain of events is not enough evidence to show that a putative cause and effect are indeed connected. Further evidence is necessary to show that each event on the chain is causally connected to the events which directly precede and succeed it. In the chapter, I show that process tracers do not yet have a consistent framework for finding such further evidence.

I then consider a counterfactualist framework for finding evidence that the links of a chain of events are genuinely causal, viz. James Woodward's interventionist theory of causation. This theory stipulates that if we cannot specify an 'intervention' for each separate link of the chain of events, then these links are not genuinely causal, and the process-tracing argument will fail. To illustrate my claims, I again consider a case study, this time Kristin Bakke's analysis of the radicalization of insurgents' tactics in the period leading up to the Second Chechen War. I show what kind of evidence Bakke needs to connect the chain of events she studies, were she to accept Woodward's interventionist theory.

Chapter 4 – Here, There, but not Everywhere: the Causal Homogeneity Condition Underlying the Use of Case Study Research to Corroborate General Theories

Having discussed aspects of descriptive comparability and transparency in chapters 2 and 3, I spend chapter 4 formulating a precise criterion of causal comparability. To do so, I first consider in what context social scientists typically try to generalize, by looking at the general hypotheses some process tracers formulate at the end of their singular case-based research. There are several existing theories of generalization in philosophy, and I compare and contrast two: one derives from Christopher Hitchcock's highly theoretical characterization of the relationship between singular and general causation, the 'causal homogeneity condition', and the other from a more pragmatic characterization of this relationship by John Dupré and Daniel Hausman, the 'average effect condition'. I specify exactly what the criterion of causal comparability would look like if we assumed either characterization, and show in what ways either are problematic for the social sciences. To illustrate my claims, I use an example from the process tracing literature, Elizabeth Wood's study of insurgency in El Salvador.

Chapter 5 – Epistemic Homogeneity: a Taxonomy

After comparing and contrasting two accounts of generalization in chapter 4, one theoretical and one pragmatic, I spend chapter 5 moving the debate forward by

finding a middle ground between the two accounts. So, while the fourth chapter presented some of the major issues for causal comparability, the fifth chapter goes some way to solving these issues. Following an older distinction by Wesley Salmon, I distinguish between *ontological* and *epistemic* comparability (what Hitchcock calls 'homogeneity'). While the former refers to whether two cases are *actually* comparable, the latter respects that in some situations, we may not know the full story and thus need to rely on whatever evidence we have of comparability. I argue that in those cases of imperfect information, an adapted form of Hitchcock's comparability criterion can be an adequate justification for generalization. To illustrate my claims, I show that Nicholas Sambanis' theory on ethnic civil wars can be critically analysed using my new 'epistemic homogeneity' account. Doing so, I draw conclusions regarding both causal comparability and the related issue of representativeness.

Chapter 6 – Thinking About Social Mechanisms

I end the thesis by discussing the 'new mechanist' theory in philosophy of biology, which describes causal mechanisms in that field; I analyse whether this theory can be fruitfully extended to describe social mechanisms as well. I first use the conclusions reached in the previous chapters to detail the similarities and differences between mechanistic reasoning in the social sciences and in biology. The main difference, I argue, is that while many of the parts of biological mechanisms (e.g. DNA replication) occur more than once, many of the parts of social mechanisms (e.g. the Fashoda crisis) occur only once. This is problematic for the 'new mechanist' theory, because this theory relies on the notion of 'regularities' to describe why causal mechanisms in biology explain. There may be regularities in biology, I argue, but there are no straightforward regularities in social science. To illustrate this limitation to the 'new mechanist' theory, I analyse my last case study of process tracing, Kenneth Schultz's study of the Fashoda Crisis.

Chapter 2

Measuring Civil War: the Causal Constraints on Systematizing Social Science Phenomena

1. Introduction

A common position in philosophy of social science holds that certain social science terms, like 'well-being', 'disability', and 'democracy', have a variety of meanings across different contexts (Alexandrova 2008, Reiss 2008a). Although these terms pick out real phenomena, the way the terms are defined depends on moral, prudential, political, and social values. In this chapter, I will discuss one constraint on this position: I will make the case that it is not just values, but also *causal inference* that constrains our definitions of such terms. I argue that although social science phenomena can and do get systematized with different opposing definitions, nevertheless which definitions we allow depends at least in part on the causal structure of the phenomena. A direct consequence of this argument is that one can make epistemic progress regarding phenomena that are defined in multiple ways.

I will illustrate my claims with a taxonomy of one particular term for which this holds, i.e. civil war. In particular, I will discuss the development of the Collier-Hoeffler theory of the causes of civil war (Collier, Hoeffler, and Rohner 2009) and its use of a specific systematized concept of civil war. I will argue that the *systematization* of the civil war phenomenon is constrained by the assumed *causal structure* of the phenomenon. This illustrates my wider argument that although the way we measure phenomena is context-dependent, any measurement for the purpose of formulating and testing a causal theory should do justice to the various causal relations the concept may stand in.

This chapter is set up as follows. In section 2, I outline the 'pluralist' position in philosophy of measurement. I present the different decisions researchers have to make when they measure an ambiguous or a fuzzy phenomenon, showing in what

ways social scientists can systematize and represent a phenomenon. I also give some attention to empirical aspects of measurement, by looking at how one can classify an individual once a systematization and representation have been settled on. This sets the scene for my argument in section 3. There, I discuss why Collier et al. systematized and represented civil war in a certain way in their theory, and in particular how these choices were constrained by causal inference. Moreover, I argue that some of the choices that Collier et al. made when they defined civil war the way they did reveal their assumptions about the nature of civil war, assumptions which in subsequent years have been challenged because they did not do justice to the causal picture.

1.1 Social ontology

This chapter focuses on the philosophical and methodological aspects of social science measurement; broadly speaking, I am concerned with the question of how one ought to systematize phenomena of interest in the social sciences, particularly when building causal theories. Ontology must play some role in answering this question: we are concerned here with real phenomena and the various causal relations they stand in. However, though I touch upon ontological questions, I will not discuss social ontology to the extent that such authors as John Searle do (Searle 2006, 2010, 2011). This means I will not discuss such topics as emergence (how the social realm is related to the natural world of particles and fields), language, consciousness, or intentionality. Specifically, I will not comment on what the ontology of social facts is, or how social facts are created. As such, my analysis will be more closely related to the recent trend in the philosophy of measurement which "treats measurement as a knowledge-producing process and attempts to analyse the sources of its reliability, rather than the metaphysical (…) conditions of its possibility" (Tal 2013, 1168).

2. The constraints on pluralism in social scientific measurement

2.1 Pluralism in philosophy of measurement

A common position in measurement methodology is that measurement takes as its input "the constellation of potentially diverse meanings associated with a given concept" (Adcock and Collier 2001, 530). The first step in measurement is taking this variety of meanings, the 'background concept', and choosing among them to form the definition of a 'systematized concept'. Some background concepts only have one meaning associated with them (e.g. 'triangle'); most social science concepts are at the other end of this scale, however. They have a wide variety of associated meanings, and accordingly they can be systematized in a variety of ways. These concepts are dubbed 'essentially contested concepts' in the methodological literature (cf. Gallie 1956). I will dub this position, that there exists a plurality of systematizations for a given phenomenon, none of which is superior to the others, 'pluralism of systematization'.

As an example of a variable background concept, Adcock and Collier give the concept 'democracy':

> [T]here is too much reliance on claims that the background concept of democracy inherently rules out one approach or the other (…). It is more productive to recognize that scholars routinely emphasize different aspects of a background concept in developing systematized concepts, each of which is potentially plausible. Rather than make sweeping claims about what the background concept 'really' means, scholars should present specific arguments, linked to the goals and context of their research, that justify their particular choices. (Adcock and Collier 2001, 532)

As another example, consider the concept 'poverty'. There are different meanings associated with poverty, all of which are plausible; absolute and relative poverty are examples of different systematizations. To do any scientific analysis of poverty, one has to choose what aspects from its constellation of meanings to emphasize. Absolute poverty refers to a fixed threshold which one must fail to meet to be

considered poor: an example is the poverty line of earning less than a dollar a day (cf. World Bank 1990). This threshold does not take into account how well off other members of one's society are (no person in the UK would be considered poor under the dollar a day threshold, for instance); it is the same no matter what a country's median income is. Relative poverty, on the other hand, refers to a standard that is set relative to one's society: an example is the UK threshold of earning less than sixty per cent of the median income. As the median income in the UK changes, some people might suddenly fall below the poverty threshold, even if their absolute income has not changed. This poverty example again illustrates that the context and goals of one's research may influence the choice of systematization.

The necessary systematization of ambiguous background concepts is not the only complication for social science measurement. A second, related aspect of pluralism of systematization is that many social science terms pick out a group of phenomena that cannot be circumscribed by a set of necessary and sufficient (boundary) conditions. This issue underlies discussions in philosophy (Crasnow forthcoming) of cluster concepts (Little 1993), Ballung concepts (Cartwright and Bradburn ms.), family resemblance concepts (Wittgenstein 1953), and (in social science methodology) of fuzzy sets (Ragin 2008) and hard versus soft data (Bulmer 2001). For such phenomena, even the systematized concept may be 'fuzzy', in the sense that it cannot be straightforwardly circumscribed.

As an example, consider the term I will investigate in more detail in the rest of this chapter: 'civil war'. In common language, civil war is not something that has definite boundaries nor, it seems, is there any one set of characteristics that all phenomena we call civil wars have in common. Rather, the different violent conflicts that get called civil wars all differ from one another. The term civil war covers a range of phenomena that are spread out over time and space and differ in intensity, motivation of the parties in conflict, duration, reasons for onset, and reasons for termination. There is, arguably, more variation within the group of conflicts we call civil war than there is between some civil wars and other forms of violence like genocide, riots, or coups (cf. Cramer 2006); as such, it may seem there

is no unique set of necessary and sufficient conditions to describe the class of phenomena that tends to get referred to as 'civil war'.

In sum, the pluralist position on systematization I have described here is twofold: firstly, many social science concepts, which I have dubbed 'ambiguous', have a variety of associated meanings; secondly, many phenomena, which I have dubbed 'fuzzy', cannot be captured by a single set of necessary and sufficient (boundary) conditions, even if there are no ambiguities in meaning. At first glance, the way to respond to these two issues may seem different – after all, if a concept only suffers from ambiguity, one may just pick one meaning suitable to the study's context and goals and continue by setting out a list of necessary and sufficient conditions. In the second case, i.e. if the phenomenon we wish to measure is 'fuzzy', analysis will be more difficult; one either needs to work with a systematization that is also fuzzy, or fix a set of boundary conditions for the phenomenon, but accept that this is an idealisation of sorts.

In practice, however, these pluralist issues may blur together. Firstly, many social science phenomena might be plagued by both issues. Secondly, while there is an interesting theoretical distinction between 'ambiguous' and 'fuzzy' concepts, in both cases the implications for the working scientist are arguably very similar. That is, they need to choose a systematization that is fit for purpose. For example, Sharon Crasnow, in her analysis of the measurement of democracy, argues that the way a concept is systematized must respect the use of the research: what is considered a good concept depends "not simply [on] accuracy or truth – but other epistemic and pragmatic virtues may be relevant. The weighting of those virtues might also vary across contexts" (Crasnow forthcoming, 10). Generally, when systematizing a phenomenon one is looking to "capture the features of the case that are relevant to the goal for which the dataset is desired" (Crasnow forthcoming, 15).

In this section, I will not go into detail on how scientists ought to go about choosing a systematisation that is fit for purpose in practice; I will leave further discussion until my case study, as it helps to use an extensive example to illustrate this.

2.2 Representing systematized concepts

So far, I have discussed two issues for systematization: ambiguity and fuzziness. I have not yet given attention to another important conceptual aspect of social scientific measurement, viz. the representation of a systematized concept. This step is discussed by Adcock and Collier as the relation between a systematized concept and its 'operationalisations', and by Cartwright and Bradburn as finding "a metrical system that appropriately represents the quality or category" (Cartwright and Bradburn ms., 1). The distinction between a systematized concept and its representation is illustrated by the case of temperature. After temperature has been systematized as kinetic energy by physicists, it still has to be represented; for instance, with the centigrade scale or with the Fahrenheit scale. This temperature example illustrates an important point about the distinction between a systematized concept and its representation: even if there is a consensus definition for a particular phenomenon, as in the case of temperature, we cannot simply deduce one and only one appropriate representation.

The importance of considering representation becomes clear when we realize that representation and systematization are mutually constraining. Generally, it may well be the case that more than one representation is appropriate for the same systematized concept. For instance, we can represent civil war in several different ways. First, we may represent civil war with a dichotomous variable. In that case we judge that either a state is not in civil war, and we assign a '0' to that state, or it is in civil war, and we assign a '1' to it. Many social scientists interested in civil war use a dichotomous variable, particularly those who wish to undertake purely quantitative, statistical research. For instance, the Collier-Hoeffler theory of civil war onset, which we will examine in section 3, looks at states for intervals of five years ('episodes') starting in 1960 and ending in 1995[1]. Those pairs of states and

---

[1] I will not discuss this temporal element to the Collier-Hoeffler theory here; it is worth noting that they assumed there was sufficient reason to treat the wars before the 1960s as significantly different and thus did not put these wars in the same class.

episodes in which a civil war broke out are assigned a '1', those in which no civil war broke out are assigned a '0'. So Rwanda in the period 1960-1965 is assigned a '1', because war started there in November 1963 according to Collier et al.'s criteria, whilst Laos from 1975 to 1980 is assigned a '0' as the conflict there ended in February 1973.

Recently, this dichotomous representation has come into question. For one, some commentators feel it is wrong to suggest that civil war is the same no matter where it takes place; the civil war in Sudan is different from the one in Colombia. But the dichotomous representation can make it seem as if all civil wars are the same by lumping the countries in which they occur together under the label '1 – engaged in civil war'. This means that the dichotomous representation does not do justice to the way these commentators have systematized the background concept civil war (even if they have not outlined this systematized concept explicitly). What is represented is a cruder concept than the intended systematized concept of civil war.

Unease about the dichotomous representation for civil war suggests researchers need to refine their systematization of civil war. This kind of back-and-forth process of mutual adjustment between how phenomena are systematized and how these systematized concepts are then represented is typical in devising measures in the social sciences. A systematized concept and its representation must get changed in tandem (Cartwright and Bradburn ms.). A change or improvement in either one of them typically produces the need for change in the other. In the remainder of this section, I will discuss two particular options for representing social science concepts: using scales, and using tables of indicators.

2.2.1 Representing concepts with scales

Stanley Smith Stevens' account of the different types in measurement (Stevens 1946, 1951) is a typology adopted by many philosophers of measurement (Luce and Suppes 2002, Suppes 1998). I will briefly illustrate this typology here to summarize some familiar types of representation that social scientists make use of.

Stevens describes four ways of representing a concept. We have already seen the first type: we may represent a concept using a *nominal* scale by assigning different numbers, letters, or names to the different units that fall under the concept. This is the kind of representation we employ when we use a dichotomous variable, as in the case of civil war where countries are divided into two groups, conventionally labelled '0' and '1'. But the numbers are just labels – we do not treat them as numbers.

Stevens' second type of representation is the *ordinal* scale. Using an ordinal scale means *ranking* the units that fall under the concept. Here the numbers do mean something. The higher the number assigned to an individual unit, the more of the quantity it possesses. So we might rank 'degree of poverty' on a scale from 1 to 10. With a merely ordinal scale the size of the differences between any two numbers does not mean anything. An ordinal scale does not imply, for instance, that the difference in the degree of poverty between individuals assigned '4' and those assigned '6' is the same as that between those assigned '8' and those assigned '10', nor that there is twice the difference in the degree of poverty for individuals falling in either of these groups as the differences between individuals assigned '1' versus '2' or '7' versus '8'.

This contrasts with Stevens' third type of scale, the *interval* scale, which both orders individual units and has equal intervals between points with equal numerical separation. The Celsius scale for temperature is an example of an interval scale: the difference between 10°C and 20°C is the same as the difference between 20°C and 30°C. However, we cannot say that 20°C is twice as hot as 10°C, since the zero point of the Celsius scale is arbitrarily defined.

Fourth, social scientists may rank the units under study on a *ratio* scale, i.e. an interval scale with a natural zero point. An example of such a scale is the Kelvin scale for temperature. We would also for instance be using a ratio scale if we assigned the label 'severity of civil war' according to the number of deaths, e.g. if we suggested that if this civil war claimed 5,000 deaths, whilst that civil war claimed 23,000, the latter is more than four times as severe as the former.

2.2.2 Representation with a table of indicators

Another way to represent social science concepts is with a table of indicators. This is generally a good representation for phenomena where a number of features matter to us but no one or two can be singled out as essential, and where it is not clear which combinations of features are better or worse than which others. The table simply lists what values the individual to be measured takes for each of the features that matter.

For instance, the Institute for Economics and Peace has devised a common set of indicators, the Global Peace Index (GPI), in order to represent national peacefulness according to three broad themes. The first theme concerns the level of safety and security in society, and includes indicators such as ones for the perceived level of criminality in society and for the number of security officers and police. The second theme bears on indicators for the extent of domestic or international conflict, such as an indicator for displaced people and the number of deaths from either internal or external conflict. The third theme deals with the degree of militarization and has indicators for military expenditure and military capability.

As with all representations, tables of indicators have advantages and disadvantages. Two related advantages stand out. First, often this kind of representation is the only comprehensive one for the phenomenon we have in mind. As I have already discussed, when the phenomenon we care about has many aspects and no clear boundaries, simpler measures end up omitting aspects and drawing boundaries that can leave individuals on the 'wrong' side.

Second, a table of indicators provides a far more detailed picture that allows us to survey the range of pluses and minuses that contribute to assigning individuals into or out of the category. We can, for instance, measure the poverty or welfare of a country on a ratio scale via GDP per capita. But the fact that many low GDP countries have had higher life expectancy at various times than would be predicted from such a measure – like Sri Lanka, Costa Rica, the Indian state of Kerala, China, some of the Soviet socialist states, or Jamaica – suggests that the individuals in these countries are not as 'poor' as this measure would make them out to be. One

standard account is that these societies provide things that individuals cannot buy for themselves, like an educational system, clean water, and health care facilities, and many individuals in the country have access to them. This is one of the reasons we might think of, say, lack of access to clean water for its citizens as part of what we mean by categorizing a society as poor. Similarly, peacefulness is not just the absence of war, and thus simply measuring civil war onset may not be enough. It also concerns for instance how safe people feel, how many of them are displaced, and how many homicides were committed. These aspects are not captured in a dichotomous or even scaled indicator for civil conflict, but they are in the Global Peace Index.

The chief disadvantage of a table of indicators is that it does not allow for a straightforward ranking of individuals, except for rare cases where one individual performs better on every indicator in the table than does another. Yet we do want to compare, both to see how things are changing in time and to see which social systems work better for reducing or enhancing the characteristics of societies that we care about. So it is not unusual to turn a table of indicators into an index number by weighting the different aspects in some way to come up with a single number[2].

The Global Peace Index is an illustration of this. The GPI is used to score countries on peacefulness, on a scale from 1 (most peaceful) to 5 (least peaceful). This score is an average of all indicators in the index, which for this purpose are individually converted to a 1-5 scale, and which are then weighted based on what the researchers deem the relative importance of each indicator. For instance, the researchers think homicides matter more to peacefulness than weapon transfers, and thus the indicator for the number of homicides weighs more heavily than the indicator for transfers of major conventional weapons (cf. Institute for Economics and Peace, 2012). Using this score for comparison, the least peaceful countries in

---

[2] Note that for a weighted average to be meaningful, we require more than ordinal (ranking) scales; we need interval-valued scales. For a discussion of this limitation in another area, environmental decision making, see Steele et al. (2009).

2012 were Sudan, Afghanistan, and Somalia (with a GPI of 3.193, 3.252, and 3.392 respectively). Iceland, Denmark, and New Zealand were the most peaceful countries (with a GPI of 1.113, 1.239, and 1.239 respectively).

As usual there is a trade-off. For the ambiguous and/or fuzzy phenomena of social science there is generally no appropriate way to weigh the various items from a table that makes good sense across all applications. The weightings are, in that sense, arbitrary. This matters because different ways of weighing will give rise to different rankings, both for the same units across time – are they improving or getting worse – and in comparing units to each other. This provides an illustration of the fact that different methods of representation generally mean that it is really different systematized concepts being represented; different weightings correspond to different phenomena.

How do we judge which representation/systematization pair is best? The pluralists' answer, as we have seen, is that this depends on the context and aims of our research. If we want a comprehensive account of just what poverty (or lack of it) is like in a particular society, a broadly inclusive concept with an associated table of indicators would generally be best. If our study's purpose is to rank people, we will need an index. And again, which is the best index, weighting which factors in which way, depends on purpose.

2.3 Measurement and classification

In the above, I have presented the different options available to researchers when they measure social science phenomena, showing in what ways social scientists can systematize and represent a phenomenon. Researchers often devise measures because they are interested in classifying particular cases. Thus, measurement has a *conceptual* component, i.e. systematizing a background concept taking into account pluralist limitations and issues of representation, and an *empirical* component, i.e. determining which individual case fits the measure and which case does not. I will

now discuss this empirical component of measurement, and illustrate how the conceptual and empirical components of measurement are related.

Consider, for example, the classification of which individuals have Asperger syndrome. This classification is important for social policy as it may determine whether or not an individual qualifies for benefits or special arrangements in the workplace. Now assume that a particular systematized concept and representation for measuring whether a person has Asperger has been settled on. Based on this, procedures might, for instance, require individuals to score a certain number on the autism spectrum disorder severity scale. The next step is empirical, viz. classification of an individual as having or not having Asperger.

The conceptual and empirical components of classification are not entirely independent: problems that come up during the empirical step, i.e. the actual classification of individuals, may require us to reconsider our conceptual work. Returning to our example, it is difficult to design systematized concepts that cover all cases we intuitively consider as instances of Asperger syndrome; our best efforts can end up making absurd rulings in many real applications. Sometimes we feel we are confronted with a case that should fit the label, and does under a different measure than the one that has been chosen, requiring us to revisit which systematization and representation to use. In that case, we may begin by carefully noting in our final work which individuals were more problematic to classify and required careful balancing. This helps other researchers to decide for themselves whether the decisions we have made are also agreeable to them, and thus whether they will want to use our classification for their own work.

This is for instance what Nicholas Sambanis has done in his 2004 measurement of civil war onset. He has posted a document online containing extensive notes on his classification of each individual conflict (Sambanis 2004b). Out of the 119 instances he has classified as civil wars, he has labelled 20 as problematic. While these 20 conflicts did not meet all of the criteria that Sambanis uses to systematize civil war, he has nevertheless decided to classify them as such. For example, Sambanis classifies the conflict in Angola between 1994 and 1999 as a civil war, noting that

this is a problematic case under his systematization. The conflict does not meet his death threshold, but he notes that the death toll occasionally comes close and thus argues "there is sufficient uncertainty regarding the death toll to include this as a borderline case" (Sambanis 2004a, 12).

In practice, a researcher may leave things at simply noting down their more problematic classification decisions, and not take the next step of going back to change the systematization of the concept they are using. Sambanis, for instance, believes that "coding rules should never be applied too strictly and that, when we are faced with a problematic case, we should err on the side of caution, i.e. we should include such cases while making it possible to identify them at the analysis stage" (Sambanis 2004b, 825). All he tries to accomplish with his publication is to give other researchers a chance to use his dataset without having to backtrack all of his decisions; he has identified them in the dataset so that other researchers can decide for themselves if they want to include them or drop them from their analysis.

In conclusion, measurement involves both conceptual and empirical issues. Once a systematization and representation have been settled upon, classifying individual cases brings up empirical concerns. When faced with cases that are difficult to classify, researchers may decide to explicitly outline why certain classifications have been made the way they have, as Sambanis does with his classification of civil war onset in certain countries. Alternatively, they may reconsider the original systematization of a phenomenon altogether to deal with cases that are problematic under the original systematization.

There is a second kind of empirical issue that may come up in measurement, namely the ease with which one can establish whether a particular case meets the criteria for inclusion under the concept. For instance, as we will see, it may be difficult to measure the death counts of civilians in a conflict, which may lead us to avoid a systematization which includes civilian death counts in its list of necessary and sufficient conditions for 'civil war'. As will become apparent when I consider conceptual and empirical aspects of my case study, I would argue such empirical

concerns are always trumped by conceptual concerns. When faced with a particular systematization that helps us meet the goal of establishing causes of the phenomenon of interest, the ease with which we can use this systematization in practice ought to be of secondary concern.

3. Civil war measurement in the Collier-Hoeffler model

In the next part of this chapter, I will consider a detailed example of measurement to show that although ambiguous and fuzzy phenomena can be systematized in different, incompatible ways, measurement is nevertheless *constrained* by causal considerations. My case study derives from civil war studies, viz. the notion of civil war used in the Collier-Hoeffler model, developed most recently by Paul Collier, Anke Hoeffler, and Dominic Rohner (Collier, Hoeffler, and Rohner 2009)[3]. I will pay particular attention to the systematization of conflicts in the source Collier et al. base their theory on, the Correlates of War project.

In what follows, I will show that the background concept 'civil war' is ambiguous, i.e. has many different associated meanings, which we have to choose between when we systematize the concept. Moreover, 'civil war' is a fuzzy phenomenon: it does not have a straightforward set of necessary and sufficient conditions, even once we disambiguate. Unfortunately, as we will see, it is not clear that Collier et al.'s aim of finding the causes of civil war is best served by the systematization and representation of civil war they use. I will consider several alternative systematizations and representations of civil war, and thereby both attempt to reconstruct some potential reasoning behind Collier et al.'s notion of civil war, and

[3] One benefit of looking at this specific example (besides that it is a well-known study in civil war research) is that the systematization and representation used by Collier et al. have much in common with the measures of civil war onset used in rival quantitative studies. For a comparison between quantitative studies, see for instance Sambanis (2004) or Sambanis (2004b).

to formulate some further considerations for systematizing and representing civil war that will help these researchers better meet their goals.

3.1 The Collier-Hoeffler model

The Collier-Hoeffler model is a theory about the causes of civil war developed by Collier and Hoeffler whilst working for the World Bank in its Development Research Group (Collier and Hoeffler 2000). Originally, Collier and Hoeffler argued that the causes of civil war fall in the category greed rather than grievance. Greed implies causes are economic; Collier and Hoeffler argued that conflict risk increases if the financial benefits of conflict outweigh the cost. They argued that actors in a conflict (e.g. the rebels) make a rational weighting of e.g. the government's ability to finance defences, the costs incurred by rebels (like the cost of recruitments), and the amount of finances rebels can collect (like by looting or tapping into primary commodity exports). So one of Collier and Hoeffler's theses is that primary commodity export is causally related to civil war outbreak, as it is an easy source of rebel finance. On the other hand, grievance involves issues of identity; it would mean civil war is caused by, for example, ethnic or religious hatreds, inequality, oppression, or vengeance. Collier and Hoeffler performed statistical research, predicting the risk of civil war onset using proxies for greed and grievance, plus a measure for civil war onset, and then analysing which proxies were correlated with risk of civil war. They concluded that greed rather than grievance causes civil war. I will come back to this conclusion, and their methodology, below.

In later work, Collier, Hoeffler, and Rohner analysed not only greed and grievance (which they now called motivations for civil war), but also feasibility (Collier, Hoeffler, and Rohner 2009). Their reason for studying feasibility is their assumption that "in most circumstances the establishment of a rebel army would be both prohibitively expensive and extremely dangerous regardless of its agenda. The relatively rare circumstances in which rebellion is financially and militarily feasible are therefore likely to constitute an important part of any explanation of civil war."

(Collier, Hoeffler, and Rohner 2009, 3) They concluded that (despite their earlier work) there is little evidence to support the idea that motivation can account for civil war risk. However, they did claim to have found evidence for the so-called 'feasibility hypothesis', which says that "where a rebellion is financially and militarily feasible it will occur" (Collier, Hoeffler, and Rohner 2009, 1), independently of whether there is some independent motivation for conflict, like greed or grievance. Whereas greed tells us about the motivations of the insurgents, feasibility is merely concerned with opportunity, that is, with whether conditions are favourable to insurgents.

3.1.1 Methodology

The aim, both in Collier and Hoeffler's earlier study as in that by Collier et al., is to develop a theory that is useful for policy makers as well as civil war scholars. Policy makers attempt to devise strategies to prevent civil war outbreak and therefore need to know what makes certain states prone to civil war and others not. Unfortunately, however, violent conflicts are highly politicized which means that the explanations offered by advocates of either side of the conflict are biased towards that side: "the public discourse is hopelessly contaminated by advocacy" (Collier, Hoeffler, and Rohner 2009, 1). This, Collier et al. say, is why they decided to study civil war through statistical analysis of global data: it abstracts from the conflict at hand and "subjects the researcher to the discipline of statistical method" (Collier, Hoeffler, and Rohner 2009, 1).

Collier and Hoeffler's as well as Collier et al.'s methodology is a logit regression analysis. I will here focus on their 2009 study as this is the most recent and because it uses a slightly different, updated dataset compared to previous work. In this study, Collier et al. use data for 208 countries in five-year periods between 1960 and 2004 (i.e. 1960-1964, 1965-1969, …, 2000-2004). 'Logit regression' is a methodology which predicts the outcome of one 'categorical' variable (in this case, a variable measuring civil war onset) based on several predictor variables. These predictor variables are proxies for (several aspects of) greed, grievance, or feasibility. For

example, they use the proportion of young men in a society as a proxy for the proportion of the population psychologically predisposed to violence and best-suited for rebel recruitment, which they argue makes rebellion more feasible (cf. Collier, Hoeffler, and Rohner 2009, 22). They then analyse what happens to the probability of civil war onset if this proportion increases. For simplicity's sake, I will not focus here on the predictor variables nor the details of the logit regression method itself. Instead, I focus only on Collier et al.'s measurement of their categorical variable, civil war onset.

3.2 Systematizing the civil war phenomenon

Collier et al.'s study is based on a source used widely in civil war studies, namely the Correlates of War project (COW). The Correlates of War project's database is one of the most common sources for data in quantitative studies of civil war. I will illustrate in this and subsequent sections that Collier et al. sometimes unhesitatingly adopt definitions and data from that project. Concerning the systematization of civil war onset, Collier et al. use COW as follows.

Civil war onset is systematized by Collier et al. based on the definition in a 2004 study by Kristian Gleditsch[4], who in turn got his definition directly from the

---

[4] They also use an alternative measure based on the systematization in a 2002 study by Nils Gleditsch and colleagues (Gleditsch et al. 2002). This alternative measure is used to check the robustness of the statistical results they derive from performing logit regression with the main measure (i.e. they perform the entire analysis again with the alternative measure and compare the results to see which correlates turn out to be significant in both analyses). I will not discuss the second measure here as this moves us too much towards a discussion of statistical methodology and away from the issue at hand, namely the systematization and representation used by Collier et al. to measure civil war onset.

Correlates of War project (then defined by Meredith Sarkees[5]). Specifically, Gleditsch's study (Gleditsch 2004) was meant to update the Correlates of War dataset of that time, which covered (intra- as well as extra- and inter-state) conflicts up to 1997. Gleditsch' new dataset amongst others included conflicts that occurred after 1997[6], but was based on the same old definition of civil war as before, i.e. he used the old COW project definition. This means that the list of conflicts that Collier et al. use in their 2009 study is ultimately based on the definition of intra-state conflict developed by Sarkees (Sarkees and Schafer 2000). This definition is therefore the focus of the remainder of this section.

In Sarkees' research notes to the COW data, she describes the evolution of COW's definition of civil wars. COW is based on a 1982 study by Melvin Small and J. David Singer, *Resort to Arms*. They classified an armed conflict as a civil war if: (1) all military action takes place in the centre of government of the state system member (and not in e.g. its colonies)[7]; (2) the national government is actively involved; (3)

---

[5] There are different versions of COW over the years. Gleditsch (and thus Collier et al.) uses version 3.0. A fourth version has already become available since.

[6] Moreover, Gleditsch argues that the COW war data contains systematic flaws because it uses an outdated definition of what constitutes a state. This is problematic because it means civil wars that occur in states not recognized as such in COW are excluded. I will not discuss this in more detail here as this primarily affects the list of conflicts before 1920 – in fact, it has no effect at all on the list of conflicts after 1960, which Collier et al. use.

[7] This requirement is included to distinguish civil wars from e.g. colonial wars. Sarkees and Schafer (2000, 126) distinguish between an empire's *metropole* (the centre of government) and its *periphery* (the colonies). For instance, during British rule in the Indian subcontinent, England was part of the metropole of the British Empire, whereas India was not. In this old taxonomy of wars, wars against territories that are *not* part of the metropole are called 'extra-systemic', as opposed to civil.

both sides put up effective resistance (measured by the ratio of fatalities of the weaker to the stronger forces); and (4) there are at least 1,000 battle deaths[8] each year[9]. A civil war was said to end if a ceasefire lasted over six months.

It is worth noting that at the time of Sarkees' analysis the COW research group argued that the 1982 coding rules "were allowing disparate conflicts to be grouped together, or conversely allowing similar conflicts to be placed in different classifications" (Sarkees and Schafer 2000, 127). They constructed a new expanded classification of wars[10]. The only consequence of this relevant for present purposes is that the COW research group changed the above-mentioned criterion (1) for civil wars to 'military action is involved' (i.e. they removed any reference to whether this action took place in the centre of government).

In the above, I have discussed the systematization of civil war, not of civil war onset. It is therefore worth noting the distinction between the two here, because as indicated Collier et al. focus specifically on civil war onset. Civil war onset refers to the time at which a conflict becomes a civil war (e.g. a month or year). In Sarkees' systematization, a civil war would begin for instance when a conflict escalates from a lower annual death count to a death count of 1,000 per year, or when the weaker

---

[8] I will come back to which deaths are included in this concept below. For a full account of the different terms for deaths in civil conflict see Lacina and Gleditsch (2005).

[9] Note that according to Nicholas Sambanis (Sambanis 2004b) there was initial confusion over whether Sarkees' 1,000 battle death threshold was annual or cumulative. This matters: for instance, the conflict in Burundi in the 1990s did not reach an annual death threshold of 1,000 but *did* reach 1,000 deaths in total (cf. Gleditsch 2004, 241, Sambanis 2004b, f8, 818-819). Sarkees herself later said that the threshold was always meant to be annual and so I will interpret it as such in the continuation of this section.

[10] A full overview of the new typology versus the old can be found in the table in Sarkees and Schafer (2000, 128).

party is responsible for 5% of the deaths for the first time. In the next section, when I discuss the representation of the systematized concept of civil war onset, I will highlight a few decisions made that impact civil war onset in particular; in this section, however, I will for the purposes of brevity ignore the distinction between civil war and civil war onset.

3.2.1 Why this systematization?

Let us look at the four criteria defined by Sarkees to see how they are used. The criterion that military action must be involved is a general requirement meant to discern wars from non-wars. The criterion that the central government must be involved is used to distinguish civil conflicts from intercommunal conflict (e.g. between two ethnic groups, neither of which represents the government). So, Sarkees (and thus, indirectly, Collier et al.) makes the assumption that civil wars have a considerably different causal dynamic than protests or intercommunal conflict; otherwise, a study into the causes of war would not distinguish civil war from these other conflicts. If a civil war scholar were to disagree with this assumption, that scholar's systematization would most likely be different. However, we rarely see this in the (quantitative) literature. Most systematizations of civil war in some way or another involve these two criteria. Compare, for instance, Sambanis' criteria for civil war (Sambanis 2004b): he only classifies an armed conflict as a civil war if the government is a principal combatant, and if there is sustained violence throughout the duration of the conflict.

Conversely, the effective resistance and death threshold criteria seem more arbitrary and are in fact more widely debated by civil war scholars, even within the (current) Correlates of War research group itself. The effective resistance criterion is used to distinguish civil wars from e.g. massacres, where the victimized party cannot put up such resistance. However, in neither Small and Singer's original 1982 study, nor in Sarkees, Gleditsch or Collier et al. do we find a defence for the specific figure of 5% (rather than, say, 4% or 6%).

There are other systematizations of civil war that use a different threshold for effective resistance; for instance, Sambanis uses a threshold of 100 deaths inflicted by the weaker side of the conflict. Neither is there a defence for using the specific threshold of 1,000 battle deaths; there are also different battle death thresholds out there, like in the Uppsala project, which includes all events with 25 battle deaths per year as armed conflicts, and classifies conflicts that result in 1,000 battle deaths in one year and at least one battle-related death in the following years as major armed conflicts.

It would be of interest to further investigate whether indeed the 5% and 1,000 figures capture what Collier et al. aim to do, viz. finding the causes of civil war, and to what degree this specific figure influences the result of their studies. In other words, it is relevant to ask (e.g. using sensitivity analysis) whether the causal conclusions that Collier et al. draw are dependent on the 5% and 1,000 figures. If the conclusions change depending on the coding rules, then it is doubly important to decide which systematization draws the most accurate causal picture.

I will come back to this causal constraint when I discuss another aspect of their coding rule, viz. the use of five-year periods rather than one-year periods. First, however, let me discuss several aspects of the effective resistance criterion and the battle death threshold. In adopting these criteria, one assumes that deaths alone can be sufficient indications of the scale of violence. Moreover, Sarkees has decided that it is best to look at an annual, absolute figure. I will turn to these two considerations now, and show that although they may be justified *empirically*, i.e. due to how easy they are to implement in practice, it is not clear they are also justified *conceptually*, i.e. are also the best choice for finding causes.

3.2.1.1 Looking at battle deaths

According to Sarkees, "the Correlates of War project uses a battle fatality threshold of 1,000 as prima facie evidence that conflict has been sustained and severe enough to classify as a war" (Sarkees and Schafer 2000, 131). This means that they assume the severity of a conflict is most adequately measured by battle deaths alone.

Before considering alternatives to the battle death threshold, consider the following. One ambiguity with using the battle death threshold is whether it means we should measure the deaths of combatants only, or include civilian deaths as well. In the intrastate conflict data, COW includes two figures: state deaths (the deaths incurred by the armed forces of the state) and total deaths (all deaths caused by the conflict, both directly and indirectly; e.g. a civilian who died in a famine that is caused by the war should be counted). The decision made in COW to have procedures for measuring both state and total deaths meshes with their systematization, which depends on knowing how many deaths are inflicted on the governmental party (to distinguish civil wars from e.g. massacres). If we are not concerned with that distinction, a total death count would have been enough.

Note that the decision to use a death count that merges both civilian and combatant deaths is not just conceptual. With this decision, COW also manages to avoid the empirical difficulty of deciding which people in a conflict count as combatants. Bethany Lacina and Nils Gleditsch point out that in civil wars, the distinction between combatants and non-combatants is often fluid (Lacina and Gleditsch 2005). If they had used a distinction between civilian deaths and combatant deaths, Collier et al. would not have respected that fluidity.

Let me now turn to alternatives to the battle death threshold. There are other costs to conflict than just the battle deaths. First of all, there are indirect deaths because of e.g. famine and epidemics[11]. Second, conflicts have non-fatal consequences: Lacina and Gleditsch for instance argue that "a complete accounting of the true human costs of conflict would include – in addition to fatalities – non-fatal injuries, disability, reduced life expectancy, sexual violence, psychological trauma, displacement, loss of property and livelihood, damage to social capital and infrastructure, environmental damage [and] destruction of cultural treasures"

---

[11] Although Sarkees seems to point to using a death count that includes indirect deaths (Sarkees and Schafer 2000) she also admits that these numbers are rarely known.

(Lacina and Gleditsch 2005, 148). Deaths alone do not capture these sufferings, and thus we may argue that the effective resistance criterion and battle death threshold alone are not sufficient to fully capture Lacina and Gleditsch's intended systematization of civil war.

In defence of her decision to use the battle death threshold, Sarkees may point out that the *conceptual* decision to take non-fatal consequences into account would lead to *empirical* difficulties later on. Measuring deaths alone is hard[12], let alone e.g. psychological trauma. Thus, the sheer difficulty of finding on-the-ground procedures for measuring non-fatal consequences may be Sarkees' defence for not taking these into account in systematizing civil war. Moreover, in response to the request to include indirect deaths, Sarkees may point out that to get an accurate death count for indirect deaths means we must judge whether certain events would

---

[12] It is rarely easy to come to a death count; the COW data often reports the figure for total deaths in a conflict as missing. In the case of the conflict in Liberia, which had a high death count, this is not problematic, but in any case where the number of state deaths is under 1,000 this does present problems. Moreover, it means that even the state deaths are at most an estimate. There are several reasons for why measuring deaths 'on the ground' is so difficult (cf. Murray et al. 2002). Common problems in measuring deaths in a conflict-stricken state include the breakdown of health information systems during conflict, like the civil registration systems that records (the cause of) deaths. Moreover, the limited information that is available may be intentionally misrepresented because of the politicisation of most conflicts. Both sides of the conflict may attempt to distort the number of casualties they have taken – and that they have caused on the other party. As civil registers become unavailable researchers have to rely on other, more indirect measures of death counts, such as demographic analyses of census data before and after conflicts (which is difficult if we want to measure whether a conflict is a civil war while it's still going on), surveys (asking whether siblings, parents, or spouses survived) or eyewitness accounts (including press reports). None of these sources can be taken to be completely accurate.

have happened if war had not started. Counting deaths due to famine as indirect deaths is only allowed if the famine would not have occurred in peacetime. Yet to establish such counterfactuals is extremely tricky, as many of the disasters like famine or epidemics have a number of causes.

However, I would argue that both when it comes to including non-fatal consequences, and when it comes to including indirect deaths, such empirical issues with measurement should come second to considering conceptual aspects. One of these conceptual aspects is the goal of measurement: in this case, describing the causes of civil war. No matter how easy it is to measure deaths alone, instead of including other consequences or indirect deaths, it is not straightforward that describing the causes of civil war is best served by Sarkees' definition, nor do Collier et al. defend this decision. I will come back to the relation between empirical and conceptual decisions below.

Sarkees has made two other decisions in her systematization of civil war, which I take to be (silently) assumed by Collier et al., who after all use this systematization. Below, I define and analyse these decisions: her choice for an annual rather than a cumulative threshold, and her choice for an absolute rather than a relative threshold.

3.2.1.2 Annual v. cumulative and absolute v. relative

One of the main disadvantages of a cumulative threshold is that it may subsume small conflicts that last for many years and slowly accumulate deaths until the total death count surpasses the threshold set. For instance, were we to use a cumulative threshold of 25 deaths, then if the IRA killed 25 UK citizens between 1950 and 2010, we would say that the UK is in civil war for that entire period. However, this would certainly not be in accordance with everyday usage of the term civil war. Thus, using an annual death threshold may indeed be the best move for Collier et al. to respect their intended systematization of civil war. Nevertheless, we may note that

the decision to use a year rather than six months or two years as the standard is to a certain extent ad hoc. I will not consider this choice in more detail here.

Now, let us consider the absolute versus relative death count. The Greco-Turkish conflict in Cyprus is excluded from data sets on civil war that use the COW delineation because of the low absolute death count in that conflict, even though it meets all other criteria for civil war. This is because Sarkees decided to use the absolute value of the death count rather than a value relative to the population size. The reason for this choice is that civil war scholars wish to capture the intensity of the violence of civil wars and thus exclude small-scale conflicts. However, in the case of Cyprus this is arguably inappropriate: one may contend that the conflict is intense despite its low death count because the population is so small. This also illustrates my earlier claims about the problematic choices one makes in classification. As anticipated there, a select number of cases, like the conflict in Cyprus, ends up on what most scholars consider the wrong side of the boundary delineating civil wars from non-civil wars (Sambanis 2004b, 821-822).

It may seem like the conceptual problem of attempting to formulate a systematized concept which includes a case we wish to call civil war, Cyprus, can easily be solved by moving from an absolute measure of deaths to a measure relative to the population size. Some civil war scholars criticize this move, however, on the basis that the relative measure has its own disadvantages that make the choice between an absolute and relative measure more complicated. Gleditsch for instance points out that that it is difficult to get reliable information about conflicts with a low death count (Gleditsch 2004), especially if there is limited media coverage and foreign reporters are thin on the ground. In such conflicts we need to rely on other sources for an estimate of the amount of conflict. Gleditsch seems to assume there are no other sources and thus that if we do not restrict our systematization to conflicts with a high absolute death toll, "we lack a reliable basis for making statements about the total amount of global conflict or comparisons between different regions of the world or time periods" (Gleditsch 2004, 233). However, these problems with the relative measure are empirical, not conceptual. It is therefore questionable whether critics of the relative measure are in the right; as stated in relation to the decision to

look at battle deaths, I would argue that conceptual decisions ought to trump what is empirically easier.

To sum up, and to illustrate how Sarkees' systematization works in practice, let us consider a particular conflict and see how we can decide that this conflict falls under Sarkees' definition. Consider the military conflict in Liberia from April 1996 until August 1996 between its government and the National Patriotic Forces. The Correlates of War database says that 3,000 deaths were incurred by the armed forces of the government, whilst other death figures (e.g. of rebels and civilians) are unknown. Despite this gap in the data, we know this conflict answers to Sarkees' criteria because we see military action, active participation by the government, a total annual death count of over 1,000, and effective resistance (after all, 3,000 deaths are incurred by the government's forces). This conflict is thus included in the lists of civil wars by Sarkees, Gleditsch and (eventually) by Collier et al. It is worth noting that using a different definition may change the classification of Liberia. For instance, Sambanis gives an analysis of the differences in classification between his own work and the Correlates of War project and shows that using his classification, there is a continued civil war in Liberia from 1992 to 1997 as opposed to only a six-month war.[13]

Again, I wish to stress here that such empirical matters should be trumped by conceptual decisions; unless the more 'practical' systematized concept still meets the goals for research, e.g. finding the causes of the phenomenon of interest, its 'practicality' is of no benefit to the researcher. Generally, when scholars try to find the causes of civil war, and thus give a list of criteria for which kinds of conflicts to include and exclude from the list of civil wars, they should keep in mind the following. Could it be harmful to their analysis if they excluded certain cases for practical reasons? Leaving out cases from a statistical study can lead to mistaken causal conclusions; it may for instance mean that certain causes are not discovered

---

[13] For details on how Sambanis reaches this alternative conclusion, see Sambanis (2004a, 133-136).

depending on whether the criteria used for including or excluding a conflict from the list of civil wars are correlated with this putative cause.

3.2.1.3 Causal constraints

So, I have shown that Sarkees' systematization of the phenomenon civil war is based on a couple of assumptions about what sets civil wars apart from other conflicts (e.g. the intensity of violence as measured in direct fatalities, the participation of the government). I have shown that these assumptions are common in the literature, but not always extensively defended; some decisions even seem arbitrary, like the choice to use the figure of 5% for effective resistance or 1,000 for the battle deaths threshold. Some of these decisions are made to avoid empirical difficulties; others have no clear motivation. Either way, I argued, such decisions ought to come second to conceptual matters, viz. the causal structure of the phenomenon.

Before I continue by looking at the representation of civil war onset, I wish to discuss one last decision made by Collier et al. that I believe best illustrates the causal constraints on systematization. Using Sarkees' four criteria for civil war means that Collier et al. make a distinction between states that are in civil war and states that are not, while they have no systematized concept to help them measure different kinds of civil wars, nor a systematized concept with which to uncover other kinds of conflicts (like, for instance, civil conflicts of a smaller scale than civil wars). To draw a sharp line between civil wars in general and all other conflicts is an important decision in itself. It means amongst others that Collier et al. do not wish to distinguish between different types of civil war, but instead treat it as a *homogeneous category*.

Political theorist Harry Eckstein (1964, 1965) discussed this specific issue. According to Eckstein, the category 'civil war' must be delimited so that some generalizations must apply to all the cases included and *only* to these cases. Classification must thus "strike a reasonable balance between inclusion and exclusion, so that both valid

generalizations about phenomena and convincing falsification are possible" (Eckstein 1964, 18). Problematically, Eckstein hoped that we might one day discover one particular feature of civil war that, when known, would mean we could derive every other characteristic of civil war as well. This hope is incompatible with the fact that civil war is associated with a constellation of properties and does not have clear boundaries.

Collier and Hoeffler discuss this decision themselves. They say that although some data sets define conflicts in terms of the underlying issues (as is the case, for example, when scholars make a distinction between ethnic and non-ethnic civil wars) they have decided not to do so because "the classification of conflicts according to their causes does not seem helpful to us if we want to analyse the causes of civil war" (Collier and Hoeffler 2001, 5).

Unfortunately, there are reasons to believe this decision is problematic. Collier et al. looked for a correlation between civil war onset and the degree of ethnic diversity of the state (meant to capture the number of different ethnic groups in that state). For some values of the threshold and criterion they find a small correlation; for others, none. This suggests that ethnic diversity is not a real cause of the kinds of conflicts we are trying to focus on using the concept civil war (Sambanis 2001).

But this is surprising because a large body of literature in international relations and a number of case studies in different countries paint a different picture. For instance, ethnic conflict in Burundi lends credence to the idea that there is a positive correlation between ethnic diversity and the incidence of civil war. In these other kinds of studies, it seems as if countries that are ethnically diverse are likely to develop conflicts. Why then does it not show up in the statistical research? The reason that some social scientists, like Sambanis, offer is the decision to group together conflicts in a category as general as civil war. By averaging across all conflicts that satisfy that systematized concept, we lose the information about ethnic diversity.

The following is evidence for this claim. In 2001, Sambanis performed a quantitative study to systematically analyse differences between the causes of ethnic and non-

ethnic civil wars (Sambanis 2001). In this study, an ethnic civil war is defined as a "war among communities (ethnicities) that are in conflict over the power relationship that exists between those communities and the state" (Sambanis 2001, 261). Using this definition, Sambanis found a significant positive correlation between the level of ethnic heterogeneity and ethnic civil war onset (and, as expected, not between ethnic heterogeneity and non-ethnic civil war onset). So, his conclusions support the idea that there is an important difference between ethnic and non-ethnic wars. This is important for (statistical) civil war studies, but also for the design of policy to e.g. prevent civil wars. If ethnic and non-ethnic wars have different causes, our policy for preventing such wars may differ also.

It is important to recognize here that, unlike what Collier et al. claimed, Sambanis did not 'classify conflict according to their causes' to then measure the degree of correlation between these classes and the causes used. The partition into *ethnic* and *non-ethnic* civil wars (wars fought over ethnicity versus other issues) is not identical to a partition according to *ethnic diversity* (the number of different ethnic groups in the state). A country can in theory have a low degree of ethnic fragmentation and still be at ethnic civil war; two ethnicities is all it takes. And, vice versa, a country can have a high degree of ethnic fragmentation and yet not be at ethnic civil war – there is no guarantee that a civil war is fought between ethnic groups, it may have started for different reasons.

In sum, the decision by Collier et al. to use a concept that is very general may make some causes invisible if we average out differing correlations for subclasses of the concept. Perhaps ethnic conflicts have different causes than non-ethnic conflicts, but these causes average out in statistical study unless we make this explicit distinction. In that case, Collier et al.'s decision to measure civil war may not be such a wise move. Sambanis' conclusion that there is a positive correlation between ethnic heterogeneity and ethnic civil war onset may therefore prove highly problematic for Collier et al.'s argument. This, then, illustrates my more general claim that causal constraints can and should influence systematization.

3.3 Representing the systematized concept

We have seen that Gleditsch and Collier et al. both base their work on Sarkees' definition of civil war. Conversely, the representation of civil war onset does take different forms in either study. I will focus exclusively on the representation by Collier et al. here and show how causal constraints influence this representation.

As I showed above, Collier et al. use data concerning 208 countries in five-year periods between 1960 and 2004 derived from Gleditsch's updated list of civil war. They represent civil war onset with a dichotomous variable. If civil war breaks out during a five-year period, they code this as a '1', and if it does not, they code this as a '0'. If civil war continues into the next five-year period, they code the variable as missing.[14] If two wars break out during one five-year period, they still only code this as a '1' (not a '2'). So, for instance, Mozambique is assigned '1' for the period 1975-1979, seeing that conflict broke out there in 1979 if we follow the COW criteria. It is assigned 'missing' for the periods from 1980-1984 until 1990-1994 because conflict only ended in 1992. It is assigned '0' for the periods after that, so 1995-1999 and 2000-2004. Rwanda, where a civil war occurred from 1990 to 1993 and a second civil war started in 1994, is assigned a '1' for the period 1990-1994.

Collier et al.'s use of a dichotomous variable to represent civil war onset in a specific state is part of their statistical methodology, logit regression analysis. This methodology still leaves them with several options, however. Though it is clear why Collier et al. associate civil wars with a specific state, I see no reason for using a five-year period rather than e.g. a one-year period or a ten-year period. In fact, Collier et al. have access to a much more specific set of dates. For example, consider the Burundian Civil War between Hutu and Tutsi tribes over control of the government in the seventies. The Correlates of War database composed on the basis of Sarkees' definition says this civil war occurred between 30 April 1972 and 25 May in 1972. Gleditsch merely says that there was a civil war in Burundi in 1972. Collier

---

[14] Their motivation for the latter is that they want to study civil war *onset*, and not civil war *duration*.

et al. in turn use an even more broad representation and assign a '1' to Burundi for the period 1970-1974.

We may ask how causal inference could improve this aspect of Collier et al.'s representation of civil war onset; would it not be better if they used a more specific set of dates given that they have access to them (i.e. seeing that it is not because of practical issues)? Collier et al. do not present a defence of this choice (causal or otherwise), and neither do Collier and Hoeffler do so in their previous papers on greed versus grievance. However, if we turn to other political scientists, we find the choice for a five-year period has been criticized on the basis of causal inference. James Fearon has constructed a version of Collier et al.'s analysis with a period of one year rather than five (Fearon 2005, 496-498). In this article, Fearon focuses particularly on the purported correlation between primary commodity exports and the risk of civil war onset; he argues that the use of a one-year period changes this correlation. Thus, Fearon argues, "the strength of the observed association [between primary commodity exports and higher risk of civil war onset] depends heavily on the choice of sample framing" (Fearon 2005, 498). Again, it seems that the representation of civil war onset cannot be devised independently of causal knowledge.

3.4 Meshing systematization and representation

In conclusion, given that Collier et al. need a quantitative indicator of civil war onset for their statistical work, there are still many options open to them. Their choice between these options is based on what meshes best with their systematized concept. I will end my analysis by presenting several alternative systematization-representation pairs.

I have already mentioned several alternative ways of representing systematized concepts when I discussed Stevens' four types of scales, and the option of using a table of indicators. Instead of using a dichotomous indicator, we may represent civil war onset for instance by using a nominal scale indicating what kind of civil war

has broken out. Jan Ångström uses such a scale; he argues against the representation of internal armed conflicts by a dichotomous variable (Ångström 2001). He argues it would be better to have indicators for different types of internal armed conflicts. If we wish to build a theory systematizing internal war, then Ångström would prefer to divide internal war into narrower concepts (e.g. ethnic conflict). His argument for this is that a theory based on internal war would be too general, because internal war covers types of conflict that have different dynamics. He points out, for example, the distinction between ethnic and ideological wars:

> [B]ecause elites in ethnic conflicts do not have to convince their followers of group belonging (…) they can focus on the relationship between territory and population and engage in acts of 'ethnic cleansing'. (…) [A]ctions [in ideological conflicts] have to be more moderate to not only maintain support from its already convinced followers but also to create support from the potential pool of support lying dormant in the state or from those who support the opposing side but can be persuaded to convert by offering more lucrative socio-economic plans for the future. Atrocities against civilians are thus less common in ideological civil wars than in ethnic civil wars. (Ångström 2001, 94)

So, in general, Ångström is afraid that if we build a theory based on the general term 'internal war', we may lose track of these distinct dynamics. Therefore, Ångström systematizes internal war as an umbrella term and thus he is not satisfied with a dichotomous indicator as his representation. He decides on a typology that distinguishes four types of internal armed conflicts (ideological conflict, leadership conflict, resource conflict, and ethnic conflict), and therefore represents internal armed conflict by a five-valued variable: a state can be in one of the four conflicts or none of these.

Alternatively, we may represent civil war onset in a broader fashion, on a scale of all kinds of conflicts ranked by intensity. Gurr and Moore use such a representation (Gurr and Moore 1997) in their attempt to estimate what variables influence the outbreak of ethno-political rebellion. They represent rebellion on a numeral scale

ranked by intensity (with 0 = no rebellion reported, 1 = political banditry, sporadic terrorism or unsuccessful coups by or on behalf of a group; all the way up to 7 = protracted civil war fought by military units with base areas). So, they are not only concerned with onset of conflict but also with the magnitude of that conflict. They differ from Collier et al. because they assume that whether a civil war or, say, political banditry happens in a state is just a matter of intensity.

This difference in assumptions is the reason that neither scaled alternative (i.e. neither a scale that distinguishes types of civil war nor a scale distinguishing types of rebellion, one of which is civil war) meshes with the systematization as used by Collier et al. I have argued that they discern civil wars as a homogeneous category that stands apart from all other types of conflict; they assume that there is a difference in dynamics between a civil war and e.g. an intercommunal conflict, a protest, or a colonial war.

We may disagree with Collier et al.'s representation, for instance because we believe that there is an important difference in dynamics between ethnic and non-ethnic conflicts (as Sambanis does in Sambanis (2001)), or no important difference in dynamics between a small-scale civil conflict and a full-blown war. This suggests that we will not find Collier et al.'s systematized concept compatible with our causal picture of conflict. Given Collier et al.'s systematized concept, their dichotomous indicator may be most suitable; this representation after all is one that groups all civil wars together and abstracts from all other particulars about that conflict (e.g. ethnic diversity). Those particulars only come in as the predictor variables. For a study into whether feasibility of civil war alone (independent of a motivation like greed or grievance) can predict civil war onset this abstraction may be best suited. I do, however, wish to emphasize, again, that such choices as the one for a five-year period should be causally constrained.

4. Conclusion

By looking at the systematized concept of civil war used by Collier et al. in their 2009 study, I have found a large set of causal assumptions that are made in this research, such as that civil war stands apart from all other conflicts, and that 'civil war' is a homogeneous category (i.e. that there is no difference in the dynamics of civil wars). Only if we give up on these assumptions can we use a completely different systematized concept. But I have also shown that we may need to revisit these assumptions, and that even the choice to study civil war onset may be mistaken if 'civil war' is too general a concept and hides causes from us in statistical study. Collier et al.'s choice to systematize civil war in general with the four criteria from Sarkees is closely linked to their choice to represent civil war onset with a dichotomous variable associated with a five-year period and a specific state. If Collier et al. chose to use an alternative representation, such as scales for conflict, they would also have to reconsider which systematization they would use.

Having considered measurement in this concrete case, let me reflect back on my more general view that social science measurement is constrained by causal considerations. In short, my analysis of the Collier-Hoeffler theory of civil war illustrates that pluralism of systematization is constrained. I have shown that the causal structure of civil war constrains our systematization of the phenomenon. After initial measurements have been taken, causal testing of a theory that uses a particular systematization may show that systematization is in fact inadequate. As I have shown, the initial belief that all civil wars have the same causes indicated to researchers Collier and Hoeffler at what level of abstraction they should systematize conflict: they measured 'civil war', rather than the more general 'war' or more specific 'ethnic civil war'. However, after further causal testing of their theory, conflict researchers like Sambanis were forced to give up the systematization 'civil war' and come up with a more specific systematization. Thus, although measurement is context-dependent (as indeed the pluralists in philosophy of measurement claim), any measurement that is used to formulate a causal theory should do justice to the causal relations the concept is in. Moreover, such conceptual decisions should trump what may be empirically 'easier'.

4.1 Generalizing these conclusions

Civil war is, arguably, not the only concept that is thus constrained. Although many social science phenomena can be systematized in different ways depending on moral, prudential, political, and social values, this does not mean that 'anything goes' in social science measurement. Social scientists' concept formation for measurement has to be causally fruitful; merely constructing a category or quantity because we are interested in it is no guarantee that we will be able to do fruitful research using this category or quantity. If we have a category that consists of individuals that have little in common except the fact that they are grouped in this category, we cannot do much with our categorization[15]. If instead it turns out that they also have other properties in common – perhaps all the states we say are in civil war are also very poor, or perhaps they all have a high unemployment rate for young men – then we have good reasons to formulate and test claims using these concepts. Because social scientists seek such 'causally fruitful' concepts, one of the important features that they look for in delineating a quantity or category is that there be some shared set of causes or some shared effects from being in the category or possessing specific values of the quantity.

So, causal fruitfulness has to guide the systematization of phenomena. It may turn out that the systematization we are initially attracted to (of democracy, poverty, or civil war) cannot be brought into the kinds of systematic relations with other concepts that allow prediction and control. There might, for instance, be no specific features that can help us to predict or explain those conflicts we want to label as civil wars, nor any interventions that would have a good chance of preventing even

---

[15] Note, however, that we should not underestimate the potential for social science groupings to start leading a life of their own after they have been formulated; a concept like *race* for instance may not have been causally fruitful initially, but the mere fact that a person is classified as belonging to a particular racial group can then impact on other factors in their life, such as educational attainment and employability.

a reasonable percentage of conflicts we want to label that way. Similarly there may not be any significant systematic effects we can expect from conflicts falling into this category. We may therefore be forced to abandon the concepts we care about and to study others that are not really the same, or are narrower. We may refine our concept of civil war, for instance, by looking at ethnic civil wars, even though we were interested in the more general concept of civil war in the first instance.

4.2 Considerations for further research

The Collier-Hoeffler model was initially developed with the aim of understanding structural characteristics of countries that experience civil war versus those that do not. As I have shown, Collier and Hoeffler for instance wanted to find out whether greed, grievance, or feasibility provides a better explanation of the risk of civil war. The measure they chose was a dichotomous indicator for civil war onset associated with a state and a five-year period between 1960 and 2004.

Although arguably Collier and Hoeffler accomplished their goal with this measure, they later found that the model was far less useful for making policy (Collier and Sambanis 2005). When the World Bank asked Collier and Hoeffler to help the Bank design strategies and programs more effectively sensitive to conflict, the old Collier-Hoeffler model proved to be too general. Policy guidelines drawn from this model were broad (e.g. on the basis of that model Collier and Hoeffler could recommend the Bank that by increasing the gross domestic product per capita in developing countries, they would be reducing the risk of civil war). To design policy, Collier and Hoeffler claimed they needed a more fine-grained model. They accomplished this by supplying the old model with several case studies from other authors. For these case studies, authors used a more broadly defined measure of conflict, which did not make a sharp distinction between civil war and other forms of violence like riots or coups. Rather, the case narratives involved a group of indicators of conflict. Collier and Hoeffler, together with the other authors of the resulting policy document *Understanding Civil War*, believed that the resulting new model accomplished their aim, saying that "it teaches us about the *process* that leads to

war, rather than focusing only on understanding structural characteristics (…). Process matters if different policy interventions can be designed to reduce risk of war at various stages of conflict." (Collier and Sambanis 2005, 19).

In the next chapter, chapter 3, I will describe and evaluate process-based accounts more generally, to see amongst others if they can provide a more accurate causal picture of phenomena like civil war.

Chapter 3

Evidence for Causal Mechanisms in the Social Sciences: Recommendations from the Manipulability Theory of Causation

1. Introduction

In a backlash against the pervasiveness of statistical methods (cf. King, Keohane, and Verba 1994), in the last decade certain social scientists have focused on finding the causal mechanisms behind observed correlations (Hall 2013, Hedström and Ylikoski 2010, Mahoney 2001, Tilly 2001). To provide evidence for such mechanisms, researchers increasingly rely on process tracing, a method which involves contrasting the observable implications of several alternative mechanisms (Bennett and Checkel 2015, Brady and Collier 2010, George and Bennett 2005).

There is a large variation in the way the term 'mechanism' is defined and used in the social sciences; for an extensive (but by no means comprehensive) overview see the table in Mahoney (2001, 579-580). Here, I wish to stress one particular distinction between causal mechanisms on the one hand and the chain of intermediate causes, or 'process', that they are thought to produce on the other. Petri Ylikoski utilizes this distinction when he speaks of mechanisms as the 'building blocks' that (interacting together in a similar way to how contributing causes interact) produce the eventual process of intervening variables or events linking a putative cause and effect[1]. Mechanisms are not linked to any particular context; the processes are what

---

[1] Arguably, political scientists Doug McAdams, Sidney Tarrow, and Charles Tilly use a similar definition of 'causal mechanism' when they claim that four of the same mechanisms (brokerage, category formation, object shift, and certification) produce a variety of events in contentious politics (including social movements, strikes, and

actually happens in a particular case. Several mechanisms can interact to produce the actual process, and the actual process will always depend on the particulars (the background conditions) of the case.

In this chapter I will mostly focus on processes, rather than the causal mechanisms themselves (those I will leave until chapter 6). It is these chains of intermediate steps between a putative cause and effect that process tracers trace. I will analyse the most recent and comprehensive set of methodological recommendations for good process tracing, viz. the set developed by Andrew Bennett and Jeffrey Checkel in *Process Tracing: from Philosophical Roots to Best Practices*. I will argue that their recommendation of simply looking for the observable implications of a mechanism is insufficient. In particular, I will argue that in order to give evidence for a causal relation between a variable X and Y, not only should one specify and find evidence for the set of intermediate variables, one should also provide evidence that each link in the chain is genuinely causal.

The process-tracing methodology literature as of yet does not commit to any particular fundamental notion of causation. Process tracing reacts to the statistical approach by arguing that finding a correlation between a potential cause and effect variable is not enough evidence for genuine causation, and that we should also investigate the intervening variables between the putative cause and effect. Process tracing, however, does not solve the problem it sets out to solve, but rather pushes the problem one step back. What, after all, is their evidence that the links between these intervening variables are cases of genuine causation?

This demand for further evidence for the links of the causal chain mirrors Stathis Psillos' distinction (Psillos 2004) between mechanistic and counterfactual causation, and his belief that mechanistic causal claims must rely on counterfactual causal claims. Psillos argues that while mechanistic approaches to causation fill in the

---

revolutions), i.e. when they argue that "similar mechanisms of change combine differently with varying environmental conditions in distinctive trajectories of historic change" (McAdam, Tarrow, and Tilly 2001, 83).

chain between a cause and effect with intermediate steps, we must provide counterfactual evidence of how these steps are causally related[2].

Here, I will use one particular counterfactual account, James Woodward's manipulability theory of causation. Woodward's manipulability theory of causation tells process tracers how to find the evidence they need: not only must process tracers study the intervening variables, but also the *intervention variables* of each link in the causal chain. Indeed, if the intervening variables do not respond appropriately to intervention, then they are not genuine intervening variables in the causal chain. After defining and giving the specific requirements for such interventions, I will show what complications arise for the social sciences. I will show what process tracing should look like in a sophisticated interventionist causal account, and argue that when we spell the method out like this we see we must be mindful of complexities like the difficulty of assessing counterfactuals, especially where causal variables are ambiguous.

This chapter then is set up as follows. First, I analyse what process tracing is and what it aims to do, giving an elaborate example from political science to support my analysis. Second, I set out the relevant aspects of Woodward's theory, in particular his notion of an intervention, including my motivation for using this theory of causation rather than another. Third, I evaluate process tracing in light of Woodward's theory, and conclude it indeed lacks evidence for genuine causation. Then, I discuss what 'interventionist' process tracing would look like, before finishing by anticipating a likely objection to this technique, the difficulty of assessing counterfactuals.

---

[2] For a similar, but less detailed argument regarding the usefulness of counterfactual reasoning for process tracing, see Lyall (2015).

2. A philosophical reconstruction of process tracing

As we have seen in chapter 1, process tracing is a mechanism-based method for analysing causal relationships. As discussed there, the term refers to two techniques (cf. Bennett and Checkel 2015), 'bottom-up' and 'top-down' process tracing. Bottom-up process tracing involves surveying a situation of interest with as little preconceptions as possible, in order to then formulate a hypothesis about possible causal connections in that situation. For instance, a researcher may spend time in a post-conflict area, interview the population to get data on their roles and interests during the conflict, and subsequently form a hypothesis about causal links between insurgency and defiance. Top-down process tracing starts with a how-possible hypothesis about the mechanisms behind some causal relation (say, the hypothesis that insurgents take up arms in conflict because of the mechanism defiance), for instance by deriving one from theory or an earlier study, and then tests this hypothesis with data collected in specific case studies (say, a case study in El Salvador, cf. Wood (2003)). Bottom-up and top-down process tracing are occasionally mixed; a researcher may start with a bottom-up study to formulate hypotheses, and continue with a top-down study to see if these hypotheses are corroborated or refuted by the evidence available. In what follows, I will look at the second type of process tracing, i.e. top-down process tracing, because I wish to evaluate how process tracers *justify* causal claims, not how they *formulate* causal claims.

First, let us consider top-down process tracing more formally, before looking at a detailed example from political science. In top-down process tracing, one formulates a hypothesis about what may be the cause of an observed effect, and by what mechanisms the two are connected. In a case study, one subsequently tries to provide support for one's own hypothesis, as well as refute any existing rival hypotheses in the literature. In the simplest case (in which there is only one hypothesized mechanism behind a chain of events), we may formalize process tracing as follows. Let us call the researcher's own hypothesis H_Z. H_Z holds that a causal mechanism Z is behind a process linking a putative cause, X, and the observed effect, Y. This mechanism has observable implications, i.e. the traces of a

set of variables Z_i such that $X \rightarrow Z_1 \rightarrow Z_2 \rightarrow ... \rightarrow Y$ (where $Z_i \rightarrow Z_j$ means that Z_i causes Z_j). It is this chain of events that process tracers trace.

The observable implications of the mechanism are generally called 'causal-process observations' (CPOs) in the literature (Brady and Collier 2004). CPOs can be thought of as "observations made while engaged in process tracing in order to evaluate a causal hypothesis, (…) observations that are salient to that hypothesis" (Crasnow 2012, 659). Methodologists Andrew Bennett and Jeffrey Checkel specify that the observable implications of mechanisms are "the facts and sequences within a case that should be true if each of the alternative hypothesized explanations of the case is true. Which actors should have known, said, and did what, and when? Who should have interacted with, worried about, or allied with whom?" (Bennett and Checkel 2015, 30). In this framework, alternative explanations cannot be true at the same time, and so finding evidence that corroborates one explanation will, at the same time, cast doubt on the alternative explanations.

2.1 Case study: diffusion of information and resources by transnational insurgents in intra-national conflict

Let us now consider an example of process tracing. Kristin Bakke, a political scientist at UCL, uses process tracing to study whether, and by what mechanisms, transnational fighters are likely to influence intra-national conflicts (Bakke 2013). In her chapter "Copying and Learning from Outsiders?", Bakke focuses particularly on the diffusion mechanisms through which transnational insurgents might transmit information and resources to local fighters.

Bakke's methods go beyond previous statistical ways of studying transnational insurgents' influence on intra-national conflict[3]. Bakke argues that a traditional

---

[3] Incidentally, we have already seen an example of such correlational analysis in chapter 2, when we discussed the (measurement of) civil war in studies like Collier and Hoeffler's.

correlational analysis of the transnational aspects of civil war "masks the mechanisms at work" (Bakke 2013, 34). Bakke's argument is that one cannot provide evidence that the connection between the presence of transnational insurgents and (one or more aspects of) violent conflict is genuinely causal by correlational analysis alone, i.e. without also specifying the chain of events connecting cause and effect. We have already encountered this criticism of correlational methods in the introduction, where I presented some of the key arguments in favour of causal mechanisms versus statistical methods. So, Bakke's challenge here is to show how it is that the presence of transnational fighters can lead to internal conflicts.

In her study, Bakke focuses on two particular mechanisms that produce this chain: relational diffusion ("the transfer of information or resources through personal networks and social bonds" (Bakke 2013, 35)), and brokerage (in which a third party "brings together information and resources from two previously unconnected parties" (Bakke 2013, 35)). I will show a more concrete example of these mechanisms below; to give just one example, Bakke hypothesizes that one of the things that diffusion by the transnational insurgents may accomplish is a shift in framing by the local insurgents, i.e. a shift in the local insurgents' view of "what they are fighting for, and who they are fighting against" (Bakke 2013, 36). Transnational insurgents might build schools or training camps, which indicates long-term mediated diffusion or brokerage. Or they might engender a direct shift in framing through short-term personal contact, i.e. via relational diffusion.

However, it is important to note that what effect relational diffusion and brokerage have on a domestic movement depends "on subsequent mechanisms – whether the new ideas resonate with the local population (…), and whether people actually adopt or adapt to the new idea or use of resources" (Bakke 2013, 36). Thus, though the same mechanisms might be present in different conflicts (e.g. transnational fighters might attempt to transfer information or resources, or bring together information and resources between two parties), the resulting chain of events or process may differ according to the 'varying environmental conditions'. If the environmental conditions are not right, e.g. because the ideas do not resonate with

the local population, it may be that although the transnational insurgents use relational diffusion and brokerage, this has no effect on the domestic movement's framing, or even the opposite effect to that intended. As Bakke puts it, "the initial diffusion mechanisms can ultimately have either helpful or harmful effects on the domestic movement's mobilization – or no effect at all" (Bakke 2013, 36).

This again reflects my earlier distinction between underlying mechanisms and the actual processes in an individual case. Depending on the background conditions in the individual case, the mechanisms can have different effects. Moreover, a different set of actual mechanisms may be at play. Thus, if we had the hypothesis that relational diffusion and brokerage were behind the shift in framing by Chechen insurgents, but found that the background conditions impeded the working of these mechanisms, we would have to conclude that it is not (just) these mechanisms that led to the observed process; we can hypothesize that at best the mechanisms are present (the insurgents attempt to use relational diffusion and brokerage) but are impeded by other mechanisms.

Finding out whether such mechanisms as those described by Bakke are indeed present is further complicated by the fact that a researcher typically cannot enter the conflict and observe the process leading to a shift in framing as it occurs. Bakke shows that researchers often can only see the first and last steps of the process that diffusion causes: researchers can for instance only investigate whether local insurgents' framing changes after the transnational insurgents enter the conflict, and whether institutions for transmitting knowledge are indeed being built, and are, moreover, built before the framing shifts.

I have mentioned above that Bakke believes one of the results of relational diffusion and brokerage is a shift in framing (i.e. a shift in what the insurgents are fighting for, and who they are fighting against). Bakke also shows that diffusion by transnational insurgents can result in a change in tactics. In particular, she focuses on transnational insurgents' "effect on a movement's use of radical tactics (…), tactics that the international community considers inappropriate wartime conduct, including the intentional killing of civilians, torture, hostage-taking, and

extrajudicial executions" (Bakke 2013, 38). Again, both mediated and relational diffusion can make a difference here: transnational insurgents can transmit "ideas about morally accepted or effective and efficient tactics" (Bakke 2013, 38) through mediated diffusion in the same schools and training camps, and "engender learning or emulation of new tactics" (Bakke 2013, 38-39). We can clearly see that environmental conditions have an impact on how successful the mechanisms of diffusion will be: learning and emulation are more likely to occur, Bakke argues, "if those tactics have proven successful elsewhere and do not contradict local norms for acceptable behaviour" (Bakke 2013, 39).

In the case of the radicalization of tactics, the observable implications researchers should look for, Bakke says, are "whether the transnational insurgents do indeed advocate or use radical tactics, as well as whether tactical innovation in the domestic movement, towards more radical tactics, takes place after the transnational insurgents enter the struggle, keeping in mind that, alternatively, the sources for innovation may come from within" (Bakke 2013, 39). In short, one of the questions Bakke is concerned with is whether transnational insurgents at least contributed to the radicalization of tactics, or whether this radicalization has an indigenous source only.

Lastly, Bakke also discusses transnational insurgents' impact on intra-national conflict through resource mobilization (i.e. the supply of "fighters, weapons, communication, know-how, and finance" (Bakke 2013, 39)), either directly (by joining forces and bringing fighters, weapons, or funds, in relational diffusion), or indirectly (by bringing the local insurgents in touch with funding sources, in mediated diffusion or brokerage). Again, the mechanisms are difficult if not impossible to observe directly, but they have observable implications in terms of e.g. an increase of resources after transnational insurgents enter the conflict, or a change in funding source.

2.1.1 Evidence for the causal hypothesis: tracing relational and mediated diffusion during the Chechen civil wars

So far, we have seen that Bakke postulates two mechanisms, relational and mediated diffusion, behind three shifts in conflict, i.e. in framing, tactical innovation, and resource mobilization, and that she does so in an attempt to go deeper into a statistical link between transnational insurgents and violent intra-national conflict. Bakke studies these phenomena in one case study, the Chechen civil wars[4]; one of her aims there, as we have seen, is to see whether transnational insurgents at least contributed to the Chechen radicalization of tactics. Bakke explicitly refers to Checkel (2008), George and Bennett (2005), and Gerring (2008b)'s

---

[4] As such, Bakke supports her claim about a mechanism that may apply to all transnational insurgents' influence on violent conflict with evidence just from the Chechen case. Sharon Crasnow has interpreted this aspect of the process tracing technique as "the causes sought in case studies are treated as singular causation", in an attempt at "seeking evidence relevant to general causal claims" (Crasnow 2012, 662). A singular causal claim describes an instantiation of a general causal claim; e.g. 'smoking causes cancer' is a general causal claim whereas 'John smoking 1 pack per day caused his cancer' is a singular causal claim. Whether this is an adequate way of interpreting process tracing, and Bakke's aim, is not immediately obvious. After all, the thing that all violent conflicts with transnational fighters will have in common is not the process (this will differ from case to case, as I have shown), but the underlying mechanisms (the way in which these transnational fighters attempt to influence e.g. radicalization). Thus, whether a philosophical account of process tracing needs to consider both singular and general causal claims, as well as the link between them, is not immediately obvious. In this chapter, I will not consider this question any further. Instead, I focus on the causal claim that Bakke makes for the Chechen case, not on whether this case study is a singular causal claim, or whether it supports any more general causal claim. I will come back to the relation between singular and general causal claims in the social sciences in chapters 4 and 5.

versions of the process-tracing technique in the methods section of her chapter. She argues that framing, tactical innovation, and resource mobilization were all influenced by transnational insurgents through a particular process, namely, one that involves both relational and mediated diffusion, and thus she tries to give evidence for the existence of the links of this process.

In what follows, I will discuss Bakke's supposed evidence for one of the shifts from the First to the Second Chechen War: tactical innovation. This part of her study is arguably the most tentative, with several alternative explanations existing in the literature and a mechanism that is impossible for outsiders to observe; as Bakke herself admits, the evidence only "hesitantly" suggests tactical innovation took place because of diffusion by transnational insurgents (Bakke 2013, 54). Thus, I believe it most clearly highlights the limitations with process tracing.

As Bakke discusses, the Second Chechen War included large-scale hostage-takings, kidnappings, and terrorist attacks, including suicide terrorism:

> [O]ne of the infamous characteristics of the Second Chechen War was a growing kidnapping-for-ransom industry, which did not to a similar extent characterize the insurgents' tactics in the first war. Another infamous characteristic associated more with the second war than the first was large-scale terrorist attacks outside Chechnya's borders, such as the Dubrovka/Nord-Ost theatre siege in Moscow in 2002 and the Beslan school siege in 2004. Moreover, in 2000, suicide terrorism became a new tool in the Chechen insurgents' repertoire of tactics. These trends suggest that the tactics of the Chechen resistance movement have changed over time, turning more radical in the sense that also civilians have become explicit targets of violence. (Bakke 2013, 41)

So, the effect (tactical innovation) is present in this case study. And so is the putative cause: as Bakke points out, there is evidence to suggest that

> [O]ver the course of the two wars, 500-700 transnational insurgents, including members of the diaspora, have fought [in Chechnya] (Moore 2007). A total of 500-700 indicates an increase over time, from the 80-90 who

were reported to be active in the first war. Some estimate there were 100-200 transnational insurgents present during the second war (B. Williams 2005b), which would suggest that the highest number of transnational insurgents entered in the interwar period and early years of the second war. (Bakke 2013, 42)

As we have seen, Bakke's tactic for providing evidence for the link between the putative cause and effect is process tracing: "the question (…) is whether this tactical innovation – towards more radical forms of action – can be traced to diffusion from transnational insurgents" (Bakke 2013, 53-54). Let me now list the evidence that Bakke presents (Bakke 2013, 54-58):

1) There is evidence for the right background conditions to make diffusion aimed at a radicalization of tactics possible, i.e. evidence that "radical tactics aimed at civilians have become (…) possibly more accepted on the part of the fighters" (Bakke 2013, 56).

2) According to Bakke, "[t]he hostage crisis that most prominently featured the influence of radical Islam was the Dubrovka/Nord-Ost theatre siege in October 2002, where the hostage takers ahead of time had made a video where they proclaimed they were seeking martyrdom in the name Allah" (Bakke 2013, 56). Before this crisis, training camps and schools had already been set up. Thus, Bakke claims "[t]iming-wise, it is plausible that the tactics were a result of learning or emulation via both relational and mediated diffusion from transnational insurgents" (Bakke 2013, 56).

3) There is some evidence for the relational and mediated diffusion of tactics *in* these training camps and schools: e.g. evidence that the most prominent transnational insurgent in the conflict, Emir Khattab (Ibn al-Khattab) posted videos of suicide bombings online; evidence that Khattab taught hostage techniques in the local training camps; and evidence that Khattab's successor, Abu Walid al-Ghamdi, called suicide attacks in Russia an effective tactic. All of these combined, Bakke argues, "suggest that the Chechens may have been exposed to and trained by transnational insurgents" (Bakke 2013, 57). (It is worth noting, however, that evidence of

this particular 'intervening factor' in the causal chain is thin on the ground: Bakke cites several secondary sources, but that is it.)

4) Last, Bakke tries to make the claim regarding diffusion more salient by a counterfactual remark: "Suicide terrorism, in contrast [to hostage taking], does not have a local historical template among the Chechens, despite centuries of conflict with central rulers. Thus in the absence of outside influence, it is unlikely that the Chechens would have turned to such a tactic." (Bakke 2013, 58)

Thus, Bakke presents evidence for the observable implications of diffusion[5]: as Bennett and Checkel put it, evidence supporting "the facts and sequences within a case that should be true" (Bennett and Checkel 2015, 30) if the mechanism was behind the chain of events. As stressed by Bennett and Checkel, Bakke does indeed hypothesize and test "Which actors should have known, said, and did what, and when? Who should have interacted with, worried about, or allied with whom?" (Bennett and Checkel 2015, 30). However, we also see that the evidence for the causal *connection* between the intermediate steps of the chain is thin on the ground.

3. Woodward's manipulability theory of causation

Now that I have presented my case study and highlighted the limitations of this study, I wish to return to my philosophical analysis of the process-tracing method.

---

[5] Besides this evidence, Bakke also considers alternative processes that could have led to the effect, and rejects most of them. To give just one example, Bakke discusses the possibility that there is a domestic alternative explanation for the hostage taking in the theatre siege, namely that the Chechens have resorted to this tactic in previous conflicts. She then rejects this explanation on the basis that there is evidence the video produced during this siege was "aimed at attracting funding from the Middle East [which] suggest[s] that transnational factors have played a role in radicalization of tactics" (Bakke 2013, 58).

In particular, I wish to highlight how one might give better evidence for the links of the causal chain. I will do so using Woodward's manipulability theory.

Let me first turn to the relevant aspects of James Woodward's manipulability theory of causation, before explicating if and how we can apply the theory to process tracing. Woodward argues that any successful description of a cause-effect relationship must refer to causal factors that can be manipulated to change the phenomenon under study. Specifically, a variable *X* is a cause of a variable *Y* if there exists some 'intervention variable' *I* which we can use to change *X*, so that *X* will then in turn change *Y* without any interference of other variables linked to *Y*. In other words, using *I* we can ascertain that *X* made the change in *Y* happen.

I have chosen to look at what would happen if the process tracer committed to Woodward's notion of causation, rather than others, for three reasons. Firstly, Woodward's theory provides an alternative to the probabilistic notions of causation that are taken for granted in the statistical approaches that methodologists like Bennett and Checkel, as well as actual process tracers like Bakke, criticize, such as the approach in the key social science methodology textbook King, Keohane, and Verba (1994). Secondly, Woodward's notion is arguably more suited to studying causal mechanisms in social science than the energy-transfer or mark-transmission notions of causation developed for causal mechanisms in areas like physics, and more applicable than accounts of mechanisms in terms of entities and activities, like Machamer, Darden, and Craver (2000); I will discuss the latter alternative in detail in chapter 6. Thirdly, Woodward's notion has not been widely applied to areas like political science and international relations, and therefore this analysis may contribute to the literature in philosophy of causation as well as to philosophy of social science.

3.1 Manipulability theory

Let me now outline Woodward's theory. The focal point of Woodward's work is his formal set of necessary and sufficient conditions for *X* to be a cause of *Y*, which form

his manipulability theory[6]. The first definition I will introduce is Woodward's definition for a direct (i.e. unmediated) cause:

> A necessary and sufficient condition for X to be a (type-level) **direct cause** of Y with respect to a variable set V is that there be a possible intervention on X that will change Y or the probability of Y when one holds fixed at some value all other variables Z_i in V. (Woodward 2003, 59)

To illustrate the use of the variable set V, consider the following toy scenario: we are interested in a Scandinavian village, asking whether, for its villagers, eating citrus fruit (X) is a direct cause of an absence of scurvy (Y). To answer that question, we cannot just feed the villagers citrus fruit for a month to see what happens to their health. We need to take into account other variables that may influence this (lack of) scurvy. So, we investigate the villagers' diet, and find out that they greatly enjoy eating liver (an excellent source of vitamin C): their liver consumption (Z) is very high. What will happen in our experiments to determine the effect of citrus consumption is the following. If we *ignore* the liver consumption, Z, of the villagers, we will find that no possible intervention on their citrus consumption, X, will change their developing scurvy or not, Y. Simply put, not eating citrus fruit will not mean that the villagers get scurvy. However, *if we keep fixed at 0* the variable Z for these villagers, we will find out that there is an intervention on X, i.e. making the villagers eat citrus fruit, that *will* change Y, i.e. whether they develop scurvy. We find that if $X = 0$, i.e. the villagers do not consume the fruit, then $Y = 1$, i.e. they develop the deficiency disease. If they *do* consume the fruit, i.e. $X = 1$, then they do not develop the disease, i.e. $Y = 0$.

The notion of a direct cause alone, however, is too basic for a complete theory of causation. Woodward calls our attention to the possibility of a variable X which influences a variable Y along some route but has no total effect on Y because X's

---

[6] The definition I will discuss here is for a general cause, not a singular cause.

influence is always cancelled out by other factors (Woodward 2003, 50)[7]. Moreover, by insisting one keeps fixed all other variables in the set V, the definition of a direct cause rules out those causes which act through a chain of variables. To deal with such cases, Woodward introduces the notion of a contributing cause:

> A necessary and sufficient condition for X to be a (type-level) **contributing cause** of Y with respect to variable set V is that:
>
> i) there be a directed path from X to Y such that each link in this path is a direct causal relationship; that is, a set of variables $Z_1, \dots, Z_n$ such that X is a direct cause of Z_1, which in turn is a direct cause of Z_2, which is a direct cause of $\dots Z_n$, which is a direct cause of Y; and that
>
> ii) there be some intervention on X that will change Y when all other variables in V that are not on this path are fixed at some value.[8]

---

[7] This issue is closely related to the notion of 'faithfulness', employed amongst others by Spirtes, Glymour, and Scheines (1993). These 'washing out' cases are cases when faithfulness, defined as being able to read all causal independence relations off probabilistic (conditional) independence, fails. The traditional example of unreliable probabilistic independence is Hesslow's birth control case (Hesslow 1976). Taking birth control pills increases a woman's risk of developing thrombosis. Pregnancy, on the other hand, is also a risk factor for thrombosis. Because birth control pills prevent pregnancy it turns out that overall, birth control pills have no effect on (or even lower) the risk of thrombosis.

[8] This second requirement is meant to sort out cases where transitivity of a causal relation fails (Woodward 2003, 58-59). To illustrate such a case, imagine that whilst having breakfast I spill coffee on my navy blazer (C), which causes me to wear a cream blazer instead ($B = c$ rather than $B = n$). Now, it turns out that at my job interview for a fashion editor position that afternoon, wearing a blazer rather than not wearing a blazer (i.e. in this scenario $B = c$ rather than $B = 0$) causes me to get the job (J). However, despite requirement (i) being satisfied (after all, there is a directed path $C \rightarrow B \rightarrow J$), we would hardly say that my spilling coffee at breakfast

> If there is only one path *P* from *X* to *Y*, or if the only alternative path from *X* to *Y* besides *P* contains no intermediate variables (i.e. is direct), then *X* is a contributing cause of *Y* as long as there is some intervention on *X* that will change the value of *Y*, for some values of the other variables in *V*. (Woodward 2003, 59)

As Woodward himself stresses, a direct cause is always a contributing cause, but a contributing cause is not always a direct cause[9].

3.2 Interventions

The notion of an 'intervention' is a crucial part of Woodward's argument. Note that there is a difference between an intervention variable and a contributing cause variable: whereas a contributing cause variable is part of the situation one is trying to analyse, the intervention variable is the means by which one undertakes this analysis. Before I discuss Woodward's rather technical definition of an intervention variable, I will introduce it with an example. I will then introduce Woodward's claim that it is not necessary for interventions to be the result of intentional human action, but that causal inferences can also be supported by properly constructed natural experiments or even by counterfactual claims about what would happen *if* an intervention on our putative cause *X* were to occur. After outlining Woodward's

---

(*C*) causes me getting the job (*J*). The causal relation is not transitive. This failure of transitivity is captured by requirement (ii): there is no intervention on my spilling coffee that will change whether I get the job. If I do not spill the coffee, I will wear my navy blazer instead. (Contrast this to the washing out cases of footnote 7; we *do* want to say that birth control pills are a cause of thrombosis, but we do *not* want to say spilling coffee causes me getting the job.)

[9] The notion of a contributing cause, working along a path rather than directly, is similar to Christopher Hitchcock's notion of a 'component effect along a causal route' (Hitchcock 2001).

theory, I will then turn to how this theory can help the process tracer give evidence for the putative causal relation between links of the process chain.

According to Woodward's theory, introducing a microfinance institution in a country will be an intervention variable I for investigating whether taking out microcredit loans (X) causes a reduction in household poverty (Y) if and only if the following things hold: first, the introduction of the microfinance institution has to increase the probability that a microcredit is taken out. Second, there must be no other factor influencing the probability that a microcredit is taken out, besides this microfinance institution (otherwise, we cannot be assured that there is not some other variable influencing the purported cause, and thereby also the effect of interest, without our knowing so). Third, and this is more difficult to ascertain in practice, the introduction of the microfinance institution should not reduce poverty in a way that is unrelated to microcredits. If it turns out, for instance, that opening a micro*savings* account also reduces households' poverty, and such accounts are offered by the microfinance institution, the third demand will fail. If the microfinance institution increases employment through staffing, and this in turn reduces households' poverty, then the third demand fails also. We would not be able to tell whether the microcredit loan or the microsavings account made the difference. In general, overlooking other ways besides X whereby I may influence Y clouds our judgement about the relation between X and Y. Fourth and last, introducing the microfinance institution must be statistically independent of all variables that reduce poverty by other means than microcredit loans. For instance, if we can only introduce the microfinance institution in regions that have a stable government, this clouds our judgement: the stability of the government could itself cause an eventual reduction in households' poverty, while the institution is getting more customers; in that case, we may be led to think it is the microcredits that influence households' poverty, while in fact it is a factor that was already present. So, we must ascertain that there are no other ways in which I can influence Y; if there were, that would mean that I gives us a misguided picture of the connection between X and Y. (To see the difference between the third and fourth requirement, consider the following. Both the third and the fourth requirement are violated if

there is a factor Z causally connected to both I and Y but not to X. Requirement 3 only captures cases in which we have $I \rightarrow Z \rightarrow Y$, whereas for requirement 4, the relation between I and Z is unknown. It may, for instance, just as well be that $I \leftarrow Z \rightarrow Y$.)

So, the four requirements in Woodward's definition of an intervention variable are:

> I is an **intervention variable** for X with respect to Y if and only if I meets the following conditions:
>
> 1) I causes X.
>
> 2) I acts as a switch for all the other variables that cause X. That is, certain values of I are such that when I attains those values, X ceases to depend on the values of other variables that cause X and instead depends only on the value taken by I.
>
> 3) Any directed path from I to Y goes through X. That is, I does not directly cause Y and is not a cause of any causes of Y that are distinct from X except, of course, for those causes of Y, if any, that are built into the $I - X - Y$ connection itself; that is, except for
>
>> a) any causes of Y that are effects of X (i.e. variables that are causally between X and Y) and
>>
>> b) any causes of Y that are between I and X and have no effect on Y independently of X.
>
> 4) I is (statistically) independent of any variable Z that causes Y and that is on a directed path that does not go through X.
>
> (Woodward 2003, 98)

In short, I is an intervention variable for X with respect to Y when we can use I to check whether X is a (direct or contributing) cause of Y, i.e. when we can use I to change X, after which X will change Y without interference from other variables causally related to Y. Using I, we will be able to ascertain that X made the change in Y happen. Thus, Woodward makes a distinction between contributing causes X, intervention variables I that we use to analyse whether a variable X is in fact a

cause, and intervening variables Z that are the means by which a contributing cause X influences its effect Y.

Woodward claims that one does not need to intervene in practice to support a causal inference. Instead, one could also look for a natural experiment, i.e. an intervention that is not the "intentional interference by a human agent" (Reiss 2005, 965). In the microfinance case, it may well be that there are two regions in the world that are similar in all crucial respects[10] except that one has microfinance institutions whereas the other does not, i.e. we would have a 'control case' and an 'experimental case'[11]. If we compared the two, taking into account all the requirements above, and found that in the country without microfinance institutions (our intervention variable) a larger proportion of households was below the poverty threshold than in the country with microfinance institutions, then this would corroborate the claim that there is a causal relation between taking out microcredits (the potential cause) and reduction of the proportion of poor households (the effect of interest).

Julian Reiss has shown that using interventions to support causal inference, whether those interventions are the result of human action or natural experiments as in the example above, is closely related to the instrumental variable technique in areas like epidemiology, empirical sociology, and econometrics (Reiss 2005). Reiss develops a sophistication of this technique to avoid some of the pitfalls with how instrumental variables are used in practice, and shows that his resulting 'causal instrumental variable technique' is "very similar to, albeit not identical with James Woodward's definition of an intervention" (Reiss 2005, 964). In the next section, I will give some account of how instrumental variables can be of use for process tracing. First, however, I wish to discuss the final form that an intervention can take.

---

[10] I concede that this requires one to specify what 'similar' would mean in this context. I will discuss this in more detail in section 4, and devote most of chapters 4 and 5 to the matter.

[11] This solution is arguably similar to the one described by Mary Morgan (2013).

According to Woodward (forthcoming), though it is sufficient to find an actual intervention variable that answers to requirements 1)-4), this is not necessary. One does not need to identify an actual intervention variable in order to establish causation; it is sufficient to establish the counterfactual of what would happen *if an intervention* on putative cause *X were to occur*.

According to Woodward, we can evaluate causal inferences based on hypothetical experiments, too. First, one would have to formulate an appropriate hypothetical experiment for the causal claim one is testing, following the definition of an intervention variable I have described above. Then, one would collect data which tells us "what the results of the (…) hypothetical experiment would be if we were to perform the experiment, although in fact we don't or can't actually perform the experiment" (Woodward forthcoming, 11).

In the next section, I will analyse if and how one could actually intervene, find a natural experiment or causal instrumental variable, or establish the result of hypothetical interventions for the putatively causal relation between the steps of a process being traced.

4. 'Interventionist' process tracing

In this section, I will apply Woodward's theory as outlined above to my example, the Bakke study of Chechen radicalization, in order to illustrate how adopting the manipulability theory would change the process-tracing technique.

At first glance, one might think that Woodward's notion of a contributing cause fits with the hypotheses in a top-down process-tracing study. In what follows I will argue that although the hypothesis H_Z has structural similarities with Woodward's notion, nevertheless the proposed methods for testing the hypotheses are quite different. In Woodward's framework, we need to show that all links $Z_i \rightarrow Z_j$ of the chain connecting X and Y are genuinely causal, which we can do using one of the three methods outlined above (actually intervening, finding a natural experiment or causal instrumental variable, or establishing the result of hypothetical interventions). Whichever method is used, we need to show that the result of a

Woodwardian intervention on each variable Z_i would be a change in the subsequent variable Z_j.

In contrast, all that the process-tracing method outlined by methodologists like Bennett and Checkel requires is that we observe the deductive implications of the purported intervening variables of the mechanism in a case study. In Bakke's analysis of whether relational and mediated diffusion by transnational insurgents caused the radicalization of tactics from the First to the Second Chechen War, it is less satisfying to simply state that schools and training camps were built between the arrival of transnational insurgents and the use of suicide bombings, than it is to clearly link that it was in those camps that local insurgents became convinced that using suicide bombings is an effective (and acceptable) tactic. That is, the schools are the actual 'difference makers'.

To contrast the two approaches in more detail, consider what further evidence Bakke needs to support the (simplified) claim that 'the presence of transnational insurgents, X, is a contributing cause of local insurgents' increased use of suicide bombings, Y, via the intervening variable of watching videos of suicide bombings that are distributed by the transnational insurgents, Z'. In this case, the manipulationist framework urges the social scientist to answer the following:

1) Is X a direct cause of Z? In other words, is there a (human, natural or hypothetical) intervention on X that will change Z or the probability of Z when one holds fixed all other variables in V at some value?

2) Is Z a direct cause of Y? In other words, is there a (human, natural or hypothetical) intervention on Z that will change Y or the probability of Y when one holds fixed all other variables in V at some value?

(In practice, as we have seen, the scientist would also investigate the observational implications of alternative mechanisms. I will not discuss this aspect here.)[12]

---

[12] These two questions only cover Woodward's requirement (i) for a contributing cause. We may ask whether we also require (ii), i.e. do we require there be some intervention on the presence of transnational insurgents that will increase the use of suicide bombings, when all variables that are not on the path $X \rightarrow Z \rightarrow Y$ are fixed? I

Using Woodward's definition of an intervention variable (Woodward 2003, 98), we can now give a list of required information. To answer question 2 above, for instance, we need to know the following:

2*) There exists a variable I_Z, or we can formulate a hypothetical variable I_Z, which

   1) causes Z;

   2) acts as a switch for Z;

   3) does not directly cause Y and does not cause any causes of Y except those on the path $I_Z \to Z \to Y$;

   4) is statistically independent of any variable A not on the path $I_Z \to Z \to Y$ that causes Y.

and analogously for question 1. Moreover, when we implement an intervention on X, we see a change in Y, indicating a causal relation between X and Y.

So, concretely, what information does Bakke need to gather? For conciseness' sake, I will focus only on question 2 here, i.e. on either finding an intervention variable I_Z,

---

would argue that we do. Like in the spilling coffee example (footnote 8) we would not call transnational insurgents' presence a cause of the increased use of suicide bombings in the following scenario: assume that the transnational insurgents only draw the Chechens' attention to videos on a different website than the one the Chechens normally visit. But if the transnationals had not been present, the Chechens would have come across the videos on the website they normally visit. In that case, even though the presence of transnational insurgents causes the Chechens to watch the videos, and watching the videos causes the increase in suicide bombings, nevertheless we would not say that the presence of transnational insurgents causes the increase in suicide bombings! There is no intervention on X that would change Y if all variables not on $X \to Y \to Z$ are fixed. A failure of transitivity must therefore be excluded, but since such cases seem rather implausible, I will not discuss transitivity in further detail here.

or establishing the counterfactual claim of what would happen if such an intervention on Z were to occur.

Let us consider what the requirements for an intervention on Z are. A variable I_Z is an intervention if, firstly, I_Z causes viewing of videos of suicide bombings; secondly, I_Z acts as a switch for the local insurgents' viewing videos of suicide bombings (i.e. makes whether the insurgents watch the videos independent of any other variables); thirdly, I_Z does not directly or through a path not on $I_Z \rightarrow Z \rightarrow Y$ cause the increased use of suicide bombings; fourthly, I_Z is statistically independent of any variable A not on the path $I_Z \rightarrow Z \rightarrow Y$ that causes the increased use of suicide bombings.

Note that simply by giving a detailed list of the evidence Woodward requires, we clearly see the connection between Woodward's interventionist framework and the process tracers' method break down. A process tracer interested in the causal connection between X and Y who follows methodologists like George, Bennett, and Checkel is not concerned with the impact of interventions. Rather, what a process tracer like Bakke does is investigate whether there are observable implications of all three factors (presence of transnational insurgents, watching of suicide bombing videos, increase in suicide bombings) present in some case study, in this case the Second Chechen War. And, as we have seen in Bakke's chapter, though she carefully collects evidence of the presence of all three factors, what she does not do in the chapter (and what she is not required to do, if we take methodological advice from George, Bennett, and Checkel seriously), is to either find an actual intervention variable or to establish the counterfactual claim of what would happen if one intervened. Thus, she will not prove that X, the presence of transnational insurgents, is a contributing cause to Y, the increase of suicide bombings, acting through a chain of intervening variables. In short, the issue is that finding observable consequences of the three factors does not show that they are causally related, because there may be alternative explanations for these observable consequences. Failure to observe the consequences can falsify, but observing them does not show that the causal hypothesis is correct.

As we have seen, there are several ways in which interventionism can help us give evidence for a causal claim like $X \rightarrow Z \rightarrow Y$. In the particular context of Bakke's work, an actual human intervention is impossible, and more generally speaking, many of the cases under study in process tracing are unlikely to be compatible with this technique.

A natural experiment would rely on finding a sufficiently similar case in which the cause was not present, to see whether the effect would be present, potentially with the use of the causal instrumental variable technique described in Reiss (2005). We would need to find a (set of) conflict(s) which are similar in every other way to the Second Chechen War, but where (to name but one step of the chain) no videos about tactics are being watched. This requires us to cash out what we mean by 'sufficiently similar'; I will come back to this below.

How about establishing the counterfactual claim of what would happen under an intervention? In that case, Bakke would need to ask: 'Could we have prevented the local insurgents from watching suicide bombing videos, in a way that is in no way connected to their use of this radical tactic through a different route? And would they have used suicide bombings less in the Second Chechen War if we had prevented them from watching suicide bombing videos in this way?'

To be concrete, in Bakke's case study she would have to work out her argument for the counterfactual claim that "suicide terrorism, in contrast [to hostage taking], does not have a local historical template among the Chechens, despite centuries of conflict with central rulers. Thus in the absence of outside influence, it is unlikely that the Chechens would have turned to such a tactic." (Bakke 2013, 58). (Note that although Bakke makes this claim, she does not go into any further detail, and as mentioned above, Checkel and Bennett never argue that such counterfactual reasoning should be part of the process-tracing technique. Moreover, this claim is a counterfactual for the causal connection between X and Y, and not a counterfactual for each intermediate step, which I have argued we need.)

Presumably, Bakke's argument for any counterfactual would have to rely on a similarity comparison, just like a natural experiment would: in the above quotation,

Bakke's assumption is that the Chechens before the arrival of the transnational insurgents are sufficiently similar to the Chechens after the arrival of the transnational insurgents, and thus that the diffusion mechanisms that the transnational insurgents set in motion are the only cause of radicalization. Therefore, if there had been no transnational insurgents, we would not have seen any use of suicide terrorism. A similarity comparison in areas like political science is, however, difficult to defend. Many of the intricate processes that are being traced seemingly occur only once, and there is reason to believe that one can always find causally relevant differences between e.g. two conflicts, or one conflict at different points in time. Thus, we might come to the sceptical conclusion that similarity assumptions will always fail.

This is an intricate issue that I will devote chapter 4 and 5 to, and thus I will set it aside for the moment. Instead, in the last part of this chapter, I will consider another objection to the case I have just made for 'interventionist' process tracing: an argument developed by Peter Spirtes and Richard Scheines (2004) that casts doubt on the possibility of finding or constructing an unambiguous intervention in social science cases.

5. The ambiguous manipulations problem[13]

In this last section of the chapter, I will discuss a particular complication with using the manipulability framework to provide evidence for the links of a causal process in the social sciences: the ambiguous manipulation problem. I will show that this problem most likely applies to generalizations in social science, but does not hold for the singular process-tracing cases under discussion in this chapter.

The ambiguous manipulation problem stems from Peter Spirtes and Richard Scheines' 2004 paper "Causal Inference of Ambiguous Manipulations". In this

---

[13] This issue was pointed out to me by James Woodward in personal communication.

paper, Spirtes and Scheines discuss the way in which their causal network theory deals with the existence of 'redundant measurements', i.e. measurements that define a variable as homogeneous while in fact it consists of two subvariables which have heterogeneous effects under manipulation. Manipulating one of the subvariables will give a different result than manipulating the other; therefore, we call such manipulations "ambiguous" (Spirtes and Scheines 2004, 834). Another way of putting this is to say that "the variable identified as causal (...) is actually a deterministic function of two [or more] underlying factors" (Spirtes and Scheines 2004, 834). In the homogeneous case, "when we intervene ideally to directly set the value of exactly one variable, *how* we set it does not matter in predicting how the rest of the system will respond" (Spirtes and Scheines 2004, 833, emphasis in original). However, when dealing with a case "[w]hen specification of the value of a variable underdetermines the values of underlying causal variables (...) we will say that manipulation of that variable is ambiguous" (Spirtes and Scheines 2004, 834).

Spirtes and Scheines use the example of the causal relation between the variable 'total cholesterol' (TC) and 'the risk of heart disease' (HD). One's total cholesterol levels are a combination of the level of high-density lipoprotein (HDL) and the level of low-density lipoprotein (LDL). This case is illustrative because the difference in causal relevance for heart disease between HDL and LDL went unrecognized for a number of years. It turns out that whilst HDL decreases the risk of heart disease, LDL increases the risk. Therefore, in Spirtes and Scheines' framework,

> Manipulation of TC is really a manipulation of HDL and LDL. However, even after an exact level of TC is specified as the target of a manipulation, there are different possible manipulations of HDL and LDL compatible with that target. (...) Hence a manipulation of TC to Medium might either lower the probability of HD (compared to the population rate), or it might raise the probability of HD. (Spirtes and Scheines 2004, 841).

Spirtes and Scheines suggest that therefore the answer to a question "What is the effect of manipulating TC to Medium on HD?" (Spirtes and Scheines 2004, 842) is "can't tell"[14].

If we apply Spirtes and Scheines' conclusions to the interventionist process tracing I have outlined above, we may conclude that one of the issues with using Woodward's manipulability framework to provide an account of the regularity between the variables of a mechanism is that in social science we are overwhelmingly faced with variables that have heterogeneous effects under manipulation. Think, for one, of the concept 'democratization'. There are many different kinds of democratization and so if, like in the LDL/HDL example, states undergoing certain types of democratization behave differently than states undergoing other types of democratization, it is highly likely that a manipulation of democratization as a total category (even under an ideal intervention) will lead to many 'can't tell's.

I will now detail why this ambiguous manipulation problem may be an issue for process tracing in order to *support general hypotheses*, e.g. as Bakke formulates for transnational insurgents' influence in intra-national conflicts in general, but *not* for process tracing in a singular case, e.g. Bakke's case study of Chechnya.

Consider again the general hypothesis about tactical innovation that Bakke is trying to support. There are two: first, that the presence of transnational insurgents causes tactical innovation through relational diffusion; second, that it does so through mediated diffusion. Yet these hypotheses make use of variables that may have subtypes, which would mean that any intervention on these variables could be instantiated in multiple heterogeneous ways. After all, there is no straightforward

---

[14] The last three pages of the paper give more detailed examples in which 'can't tell' is the answer for some manipulations, and in which the number of 'can't tell' answers increases when the possibility of having ambiguous manipulations is admitted into the model. This is also expanded upon in an earlier technical report with the same title, Spirtes and Scheines (2003).

definition of 'transnational insurgents'. It may well be that (under the same background conditions) some transnational insurgents' diffusion leads to the radicalization of tactics, whilst others' diffusion impedes this effect (just like LDL increases the risk of heart disease whilst HDL lowers the risk). Therefore, an intervention on the presence of the 'total level' of insurgents is ambiguous; when asked what their presence will mean for (radical) tactical innovation, the answer will be 'can't tell'. The problem, here, is the overgeneralization of the potential cause[15].

But what about the very specific claims that Bakke makes about the presence of Islamist fighters like Emir Khattab and Abu Walid al-Ghamdi in Chechnya? Here, I would argue, the ambiguous manipulation argument does not hold. Asking counterfactual questions like 'would the local Chechen fighters have used suicide bombings if they had not watched videos about this tactic?' is essential, and for the reasons I have outlined above, very difficult; yet it does not suffer from the criticism that the variables have heterogeneous effects under manipulations. Working out a counterfactual remark like "in the absence of outside influence, it is unlikely that the Chechens would have turned to [suicide terrorism]" (Bakke 2013, 58), based on the comparison 'suicide terrorism was not used by the Chechens in previous violent conflicts', faces the similarity problem (are the Chechens then 'similar enough' to the Chechens now to make this claim), but not the ambiguous manipulation problem (we know very firmly which transnational insurgents we are trying to intervene upon, and there are good reasons to believe that they are working towards the same effect).

---

[15] This is a different problem than the issue that the set of *insurgencies* might behave differently under the presence of transnational insurgents. Even if all transnational insurgents acted in similar ways to bring about relational and mediated diffusion, it may still be the case that some insurgencies radicalize under the transnationals' influence, while others do not.

6. Conclusion

In this chapter, I have shown that process tracers generally postulate causal hypotheses which relate a putative cause X and observed effect Y by a mechanism, Z, which has a chain of intervening variables $Z_1 \ldots Z_n$ as its observable implications. Process tracers then find a case study in which Y is present, and investigate whether $X, Z_1, \ldots, Z_n$ are also present.

Woodward holds that X is a *contributing* cause of Y with respect to V if and only if there was a set of intervening variables $Z_1, \ldots, Z_n$ such that X is a direct cause of Z_1, which in turn is a direct cause of Z_2, which is a direct cause of … Z_n, which is a direct cause of Y. Moreover, Y depends on X. As it stands, process tracing does not establish the complete right hand side of this if and only if statement. Process tracers show that a set of purported intervening variables exists, but they do not show that each link of the chain is a relation of direct causation.

I have argued that if process tracers were to commit to Woodward's notion of causation, they would have to provide evidence that there is a possible intervention to show that the relations they hypothesize are genuinely causal. I have shown that one way of gaining knowledge of such intervention variables is by comparing and contrasting one case to another, either with a natural or hypothetical experiment, which in both cases requires a sophisticated analysis of to what extent such cases are similar. I have also shown that one criticism of the applicability of my account, the ambiguous manipulation problem, arguably holds for general process hypotheses, but not for detailed singular hypotheses. In the next chapters, I will turn to the 'similarity condition' in more detail.

Chapter 4

Here, There, but not Everywhere: the Causal Homogeneity Condition Underlying the Use of Case Study Research to Corroborate General Theories

1. Introduction

As we have seen in chapter 3, social scientists have recently become more reliant on case study research to provide evidence for causal claims about the mechanisms connecting a putative cause and effect (cf. George and Bennett 2005, Hall 2013, Hedström and Ylikoski 2010, Mahoney 2001, Tilly 2001). These researchers commonly use case study results to directly infer a causal theory that applies to a range of cases[1]. However, the conditions under which such generalization is warranted have received only limited attention. In this chapter, I argue that we can interpret these researchers' attempt at generalization as an attempt to use singular causal claims to establish a general causal claim. I present a formal definition of general causal claims, show why it is difficult to meet the requirements for this definition in social science, and present an alternative, weaker definition that does not face these difficulties.

Singular causal claims assert that a relation holds between individual events, whereas general causal claims assert that a relation holds between event types. A general causal claim effectively states that all singular instances within the domain are similar in their response to the cause variable. One way in which philosophers

---

[1] Here we define a case loosely as an instance of a "phenomenon of scientific interest", e.g. a revolution, a democratic regime (George and Bennett 2005, 17).

have defined the relationship between singular and general causal claims is to say that in order to move from singular causal claims (indexed to a specific case) to a general causal claim (which makes no reference to any specific case, but only to the putative cause and effect variable), we need to show that the cases over which we generalize are 'sufficiently similar' to one another (cf. Hitchcock 1995, 280-283, Holland 1986). In brief, we need to know that in every case, the effect variable will respond similarly to the causal variable, all other things being equal. I will interpret this 'similarity' using the 'Hitchcock causal homogeneity condition', and spell it out using Hitchcock (1995)'s probabilistic framework for causation.

In what follows, I will show that in light of the problem of external validity for the social sciences, it is difficult, if not impossible, to meet the standards for Christopher Hitchcock's causal homogeneity condition; the causal relations we find in case studies are particular to certain contexts. It seems therefore that Hitchcock has set the bar too high for general causal claims in the social sciences. We are rarely warranted in inferring a Hitchcockian general causal claim from a singular case study.

In the last section of the chapter, I anticipate a different characterization of general causal claims, Hausman's average effect condition. I will show that the social sciences can more easily meet the standards for this characterization of general causal claims. However, as it will turn out, Hausman's characterization has other drawbacks. Hausman's general claims are considerably weaker than Hitchcock's general claims, and can therefore be misleading.

In the next section of this paper, section 2, I demonstrate that social scientists indeed attempt to use singular causal claims to directly infer general causal claims. Then, in section 3, I outline a probabilistic framework for such generalizations, including the Hitchcock causal homogeneity condition, and anticipate objections to this framework. In section 4, I apply this probabilistic framework to the social sciences and use this to outline the problems for generalization that are specific to the social sciences. Finally, in section 5, I present Hausman's alternative characterization of general causal claims.

2. Process tracing and its aims

Though the difficulties with using singular causal claims to corroborate general causal claims apply more broadly, I will focus my attention on process tracing here. As outlined in chapter 3, process tracing is a mechanism-based method for analysing causal relationships. In this thesis I focus on top-down process tracing (cf. Bennett and Checkel 2015), which starts with a how-possible hypothesis about what may be the cause of an observed effect, and by what mechanisms the two are connected (taking these mechanisms from, for instance, theory or an earlier study). In a case study, one subsequently tries to provide support for one's own hypothesis, as well as refute any existing rival hypotheses in the literature.

In chapter 3, I formalized process tracing as follows: let us call the researcher's own hypothesis H_Z. H_Z holds that a causal mechanism Z is behind a process linking a putative cause, X, and the observed effect, Y. In any particular case, mechanism Z has observable implications ('causal-process observations', CPOs), i.e. the traces of a set of variables Z_i such that $X \rightarrow Z_1 \rightarrow Z_2 \rightarrow ... \rightarrow Y$ (where $Z_i \rightarrow Z_j$ means that Z_i causes Z_j). It is this chain of events that process tracers trace to see whether H_Z holds for a particular case. In chapter 3, I gave the example of Bakke's study of transnational insurgents' influence on violent conflict. Bakke hypothesized that there are several diffusion mechanisms behind this causal relation, but only looked at the process these mechanisms produced in a single case, the period leading up to the Second Chechen War.

2.1 The tension between process tracing's singular and general causal claims

Sharon Crasnow has argued that process tracers use singular case study evidence to corroborate a general causal hypothesis (cf. Crasnow 2012). In particular, she postulates that "the causes sought in case studies are treated as singular causation", in an attempt at "seeking evidence relevant to general causal claims" (Crasnow 2012, 662). Crasnow problematizes this, arguing that there is no reason to think the

evidence produced by applying process tracing in a singular case study supports these general causal claims (Crasnow 2012, 665).

This is problematic because top-down process tracers work towards directly inferring general theories on the basis of their case study research. In other words, the singular causal claims resulting from case studies are thought to help us directly infer general causal claims about a range of cases. Methodologist Peter Hall, for instance, writes that top-down process tracing is "attentive to the general causal inferences that can be drawn from observing the sequence and timing of events and contemporary interpretations of those events" (Hall 2013, 22). Bennett and Checkel also say that some authors "focus their process tracing as much on explaining an important historical case (…) as on developing and testing general theories" (Bennett and Checkel 2015, 9).

However, under what assumptions ('scope conditions') one can use an observed process in a case study, and the associated hypothesis about the underlying causal mechanism, to infer an appropriately general claim about the causal mechanism for a wider population is unclear in the methodological literature. In "Process Tracing: from Philosophical Roots to Best Practices", Bennett and Checkel only briefly recognize the issue when they argue that "because causal mechanisms are operationalized in specific cases, and process tracing is a within-case method of analysis, generalization can be problematic" (Bennett and Checkel 2015, 13). Bennett and Checkel believe that one cannot figure out the scope conditions of a mechanism before the case study has been studied, because "a researcher cannot have a very clear idea whether, how, and to which populations an explanation of a case might generalize until they have a clear theory about the workings of the mechanisms involved in the case" (Bennett and Checkel 2015, 13). However, they do not provide a framework for generalization after the case study has been done.

In what follows, I consider one candidate for such a framework following Hitchcock (1995). Doing so will make clear what challenges top-down process tracing faces, and point some way towards meeting these challenges. The novel contribution of this chapter consists of two parts: first, though Hitchcock's mathematized

expression already exists, it has not been used to characterize a way in which case study researchers may formulate general claims. Second, by mathematizing generalization and its challenges, we can start looking for solutions to this issue. The mathematization will also allow me, in chapter 5, to characterize general causal claims in a way that is substantially different from Hitchcock's, but more appealing in the social sciences. I will call this alternative the 'epistemic homogeneity condition'.

2.2 Illustration: singular and general causal claims in Elizabeth Wood's study of mobilization

To illustrate what evidence process-tracing researchers need to test general causal claims on the basis of singular case study evidence, consider an example of process tracing that is hailed by methodologists like Jeffrey Checkel as a study that "sets the standard for how to capture the presence and role of causal mechanisms" (Checkel 2013, 19), and that has won prizes from both the American Political Science Association and the American Sociological Association. Yale University political scientist Elizabeth Wood used process tracing to study the mobilization of rural people in the Usulután region of El Salvador during the Salvadoran Civil War (Wood 2003). She spent fifteen years in the region doing ethnographic fieldwork, collecting oral histories, and trying to exclude alternative explanations of mobilization put forward by previous researchers.

Wood's main hypothesis is the singular causal claim that during the Salvadoran Civil War, rural people mobilized into the Frente Farabundo Martí para la Liberación Nacional (FMLN) rebel forces because of three mechanisms[2]: because

---

[2] Note that Wood herself does not use the term 'mechanism'; this is an interpretation by authors like Jeffrey Checkel. Instead, Wood refers to the three factors as 'reasons' or 'reasons for acting'. For a short discussion of this term and why she chose it, see Wood (2003, 231).

they came to value participation per se, because of 'defiance' (a refusal to acquiesce), and because of 'pleasure of agency' (the "positive affect associated with self-determination, autonomy, self-esteem, efficacy, and pride that come from the successful assertion of intention" (Wood 2003, 235)). Wood contrasts these three mechanisms with alternative explanations of mobilization, such as the hypothesis that mobilization is mainly caused by peasants' consideration of the potential material benefits of joining the rebel forces (making peasants mobilize because of e.g. their 'desire for land').

Though the main hypothesis in Wood's work is a singular causal claim about Usulután, in the final chapter of her book, Wood moves beyond claims about Usulután, and postulates hypotheses about insurgent collective action in general based on her findings. Wood does not seek to test these general causal claims. I will use Wood's study to illustrate what further evidence would be necessary to infer the general causal claims she hypothesizes. Wood writes:

> The reasons for insurgent collective actions stressed here – participation, defiance, and pleasure in agency [*sic*], as well as closely related reasons, such as self-respect, honor, dignity, recognition, and reputation – appear to have played powerful roles in other, quite diverse, cases of collective action by long subordinate social actors. (Wood 2003, 246)

Amongst these other cases, Wood claims, are the US civil rights movement, 1940s Lithuania, and 1970s Peru. According to Wood,

> Further research is needed to trace more precisely the role of (…) reasons for [collective] actions and how they emerged in particular social and cultural contexts. *If the argument presented here holds more generally, as I believe it may*, such tracing would find that participation in insurgent activities has cultural as well as political consequences that may, in some circumstances, reinforce insurgent values and beliefs such that insurgency continues despite high risks. (Wood 2003, 251, emphasis added)

In other words, Wood takes her case study in El Salvador to suggest a *general* causal hypothesis (about the class of events 'insurgencies'), i.e. the claim that insurgent

collective action is caused by participation, defiance, and pleasure of agency, and not (just) by material interests.[3] As stated above, Wood is cautious about the scope of the generalization she makes, and she does suggest further evidence is needed even within that scope. She does not attempt to find such further evidence in her study.

3. The causal homogeneity condition

In what follows, I will zoom out from the particulars of Wood's study, and frame the relation between singular and general causal claims using a probabilistic framework that follows Hitchcock (1995). In particular, I utilize the 'Hitchcock causal homogeneity condition' that is meant to characterize the relation between singular causal claims and general causal claims. Then, in the next section, I return to the particulars of social science, analysing whether the demands for the Hitchcock causal homogeneity condition can be met, by studying what evidence one needs to make a Hitchcockian generalization of a singular causal claim like 'defiance caused the Usulután peasants to join the FMLN' to a claim like 'defiance causes civilian mobilization'.

There are some limitations to using the Hitchcock causal homogeneity condition here as a tool for better understanding the inferential relation between singular case

---

[3] Besides this potential generalization from the Salvadoran case to other insurgencies, there is another generalization inherent in Wood's study. Wood has interviewed 200 *campesinos* (peasants), chosen through *campesino* organizations, and only in parts of El Salvador (four sites in Usulután and one in the municipality of Tenancingo) that she believes were representative for the rest of the insurgency (Wood 2003, 52-54). To go from such causal-process observations of selected insurgents' motivations to claims about the motivations of all insurgents in the conflict is also a generalization. So, there are several points in case study research when researchers can be said to generalize from individual events to event types.

study claims and general hypotheses. Hitchcock, we must realize, characterizes the relationship between singular and general causation. As such, his causal homogeneity condition is a definition, which tells us that when the conditional probabilities of all individual cases are the same, we can make a general causal claim that applies to this set of individuals. However, the Hitchcock causal homogeneity condition does not tell us how to work out whether we should *expect* the conditional probabilities of all individual cases to be the same in a particular case. I will come back to this issue below.

3.1 Singular and general causation in a probabilistic framework

As indicated above, singular causal claims assert that a relation holds between events, whereas general causal claims assert that a relation holds between event types. Christopher Hitchcock famously uses probabilistic causation to characterize the distinction between singular and general causal claims (Hitchcock 1995)[4]. I will follow his account here.

---

[4] There are several positions on the relation between singular and general causation in the literature. Hitchcock (1995) responds to earlier theories about this relation by amongst other Ellery Eells, who (on the basis of his particular probabilistic framework for singular and general causal claims) concluded that singular and general causal claims describe two distinct kinds of causal relation, and that "(1) very little (if anything) about what happens on the token [i.e. singular] level can be inferred from type-level [i.e. general] probabilistic causal claims, and (…) (2) very little (if anything) about type-level probabilistic causal relations can be inferred from token-level probabilistic causal claims" (Eells 1991, 6). Hitchcock (1995) uses his own refinement of the probabilistic theory of causation (first published two years earlier, see Hitchcock (1993)) to introduce a new characterization of the relationship between singular and general causal claims that goes against Eells' claims. Hitchcock shows that his own characterization of the relation does not fall

In probabilistic causation, all causal claims describe a conditional probability function. The putative cause either increases or decreases the probability of its effect (in the case of promoting or inhibiting causes, respectively), as long as we keep any background conditions that potentially influence the effect fixed at some value.

To introduce the framework in Hitchcock (1995), consider his example of John, who smokes one pack of cigarettes per day and contracts lung cancer. We wish to evaluate whether John's smoking one pack per day, as opposed to none, has caused his lung cancer. To model such singular causal claims, Hitchcock constructs a probability space for the individual i of interest[5] (e.g. John), consisting (as is typical in probability theory) of a sample space Ω_i, a set of events $\mathcal{F}_i$, and a probability function $P_i: \mathcal{F}_i \to [0,1]$. The sample space Ω_i consists of all possible instances of the process being modelled. In our smoking example, Ω_{John} contains the outcomes $X_{John} = 0$, $X_{John} = 1$, $X_{John} = 2$, and so on, denoting that John does not smoke, smokes one pack per day, two packs per day, and so on. The set[6] of events $\mathcal{F}_i$ contains combinations of outcomes from Ω_i (we call these combinations 'events'; i.e. each event is a subset of Ω_i containing zero or more outcomes; an example is 'smoking more than one pack per day'). Hitchcock stresses that one can partition Ω_i into background contexts $G_i^1, G_i^2, ...$, so that we can assess the causal relationship against background conditions that possibly influence the effect we are interested in (e.g. living in a polluted city, being in regular contact with asbestos).

---

foul of the counterexamples Eells presented. Hitchcock thereby concludes that under his framework for characterizing the relationship between singular and general causation, the two are *not* independent. I will not discuss this history of Hitchcock's account any further here.

[5] What we consider to be an 'individual' depends on the process we are trying to model. So, an 'individual' can e.g. be a particular die, a particular coin, or a particular person (in the case of John's lung cancer).

[6] Strictly speaking, $\mathcal{F}_i$ is a sigma algebra (cf. Hitchcock 1995, 259, 286 f2).

Given an individual i and background context G_i^α, we can define a distribution function $f_{i\alpha}$ (dubbed the 'little probability space' by Hitchcock). The domain of $f_{i\alpha}$ is the set $\{x\}$ of possible values of the causal variable (e.g. the set $\{0,1,2,...\}$ of the number of packs smoked per day by John). The range of $f_{i\alpha}$ is the set of probability functions $P_i(E_i|X_i = x, G_i^\alpha)$, which tell us the probabilities of the effect E_i given that the causal variable X_i has value x and given background context G_i^α. In short, $f_{i\alpha}:\{x\} \to P_i$ such that $f_{i\alpha}(x) = P_i(E_i|X_i = x, G_i^\alpha)$. We claim, then, that John's smoking one pack per day as opposed to no packs per day caused his lung cancer if[7]
$$f_{John,\alpha}(1) > f_{John,\alpha}(0).$$

3.2 Causal homogeneity

So far we have considered singular causal claims only. Every 'mathematical entity' in the framework above is indexed to an individual case (e.g. John in the smoking example). Hitchcock characterizes the relationship between singular causal claims, indexed to the individual, and general causal claims, which make no reference to particular individuals, but only to the putative cause and effect variable, in terms of whether the individuals over which this general claim ranges are 'sufficiently similar' (cf. Hitchcock 1995, 280-283). Consider, for example, the general causal claim that 'smoking causes cancer' for a set of individuals including both John and Mary. Hitchcock interprets this general causal claim as saying that the relation between smoking and cancer is sufficiently similar for John and Mary. In brief, he argues that when we know that every individual will respond similarly to the causal variable being at a certain value, all other things being equal, then we can

---

[7] Hitchcock himself (cf. Hitchcock 1995, 159-160) is ambiguous about whether this relation needs to hold for *a sufficient number of* background contexts $G_i^\alpha, G_i^\beta, ...$ or for *all* background contexts ('contextual unanimity'). Like Hitchcock, I will not discuss this aspect of probabilistic causation here. For a further discussion of this topic see e.g. Cartwright (1979), Dupré (1984), and Eells (1991).

make a general causal claim. John and Mary, all other things being equal, are *equally likely* to develop lung cancer when they smoke a certain number of packs per day. This characterization of the relation between singular and general causal claims is what I will call the 'Hitchcock causal homogeneity condition'.

Let us make the Hitchcock causal homogeneity condition a little more concrete using the framework above. We have seen that for an individual i and a background context G_i^α, the little probability space is $f_{i\alpha} = P_i(E_i|X_i = x, G_i^\alpha)$. Now, the Hitchcock causal homogeneity condition, applied to each context G_i^α, is for $f_{i\alpha}$ to be independent of i. In other words, if for all $i, j \in I$, the set of individuals about which we make the general causal claim, we have $f_{i\alpha} = f_{j\alpha}$, then we can make a general causal claim about this set of individuals I. This will allow us to construct a new probability space $\{\Omega', \mathcal{F}', P'\}$ based around a probability function $P'(E|X = x, G^\alpha) = P_i(E_i|X_i = x, G_i^\alpha)$ for all $i \in I$. This function P' is effectively a general causal claim.

Note that, as stated earlier, the Hitchcock causal homogeneity condition is a definition of generalization, not a method for finding such generalizations. In subsequent sections, I will use the definition to specify what evidence is required, in the social sciences, to support a Hitchcockian general causal claim, and I consider several difficulties for collecting such evidence. First, however, I end this section by noting a similarity between Hitchcock's definition and another definition common in the literature, the 'unit homogeneity condition'.

3.2.1 Holland's unit homogeneity assumption

The Hitchcock causal homogeneity condition as I have outlined it above is similar to the 'unit homogeneity assumption' formulated originally by Paul Holland (1986), and outlined also in King, Keohane, and Verba (1994); see also the analysis in Psillos (2004). The unit homogeneity assumption centres on two kinds of variables: a 'potential cause variable', S, which can influence units u in a population U, and a 'response variable', Y, which represents the effect on a unit for each value of S. Say S

consists of two variables, viz. t (treatment) and c (control). The 'causal effect' of the treatment t on a unit u as measured by response Y and relative to the control c is defined as the difference between $Y_t(u)$ and $Y_c(u)$ (Holland 1986, 947).

As Holland points out, the problem with this definition of the causal effect is that we can only measure $Y_t(u)$ or $Y_c(u)$ for a unit u, and not both. So, we cannot observe the causal effect for a single unit u. The (practical) solution to this problem, Holland argues, is to derive the value of one of the variables from an experiment on a similar unit u'. This requires a 'unit homogeneity assumption': the assumption that the two units u and u' will respond in the same way to the treatment relative to the control, i.e. that $Y_t(u) = Y_t(u')$ and $Y_c(u) = Y_c(u')$ (Holland 1986, 948).[8] Holland thus argues that, in order to do the inferential work that he requires, we have two units u and u' for which a Hitchcockian general causal claim regarding treatment and control can be made.

4. The causal homogeneity condition in social science

Now that I have discussed Hitchcock's characterization of the relationship between singular and general causal claims in terms of the causal homogeneity condition, I wish to return to how we can characterize the relation between singular and general causal claims in the social sciences in particular. In this section, I will first discuss what evidence one needs to support general claims as characterized by the Hitchcock causal homogeneity assumption in the social sciences. I will begin by stating a simple version of how one would give evidence for the condition in

---

[8] A special version of the unit homogeneity condition is the 'temporal stability condition': there, one does not derive the value of one of the variables $Y_t(u)$ and $Y_c(u)$ from the value of another variable u', but rather from the value of the same variable at a later time, which means we must assume "the constancy of response over time" (Holland 1986, 948).

Wood's study, before giving a general overview of the complications of giving evidence for Hitchcock's causal homogeneity condition in the social sciences.

4.1 Causal homogeneity in process tracing

Let us return to the particulars of social science, analysing what evidence one needs for the Hitchcock causal homogeneity condition in the context of connecting a singular causal claim like 'defiance caused the Usulután peasants to join the FMLN' to corroborate a general claim like 'defiance causes civilian mobilization'. A social scientist who accepts the Hitchcock causal homogeneity condition as a useful characterization of general causal claims, will have to detail this condition in the manner outlined below.

Before we can generalize, we must construct the probability space $\{\Omega_S, \mathcal{F}_S, P_S\}$, with S the case of the Salvadoran Civil War. The causal claim that in El Salvador, the positive affect associated with self-determination in a rural population (the 'pleasure of agency' these people feel), $D_S = 1$ (rather than $D_S = 0$), causes mobilization of this same rural population, M_S, is a claim about the little probability space $f_S(x) = P_S(M_S|D_S = 1, G_S^\alpha)$. The claim is that $P_S(M_S|D_S = 1, G_S^\alpha) > P_S(M_S, G_S^\alpha)$[9].

The second step is the generalization. We should construct, firstly, the larger set I over which we wish to generalize, and which contains our case study S. Wood herself, as we have seen, hypothesized that this set might contain C, the US civil rights movement; L, Lithuania in the 1940s; and P, Peru in the 1970s. Assuming for the moment that this is an exhaustive list, we might therefore construct $I = \{S, C, L, P\}$. Secondly, to give evidence for the Hitchcock causal homogeneity condition, one ought to show that for all $i, j \in I$, $f_{i\alpha} = f_{j\alpha}$. Thus, for example, we want to know whether $P_S(M_S|D_S = 1, G_S^\alpha) = P_C(M_C|D_C = 1, G_C^\alpha)$. In other words, if

---

[9] By carefully delineating G_S^α, we can do justice to the claim that defiance causes mobilization *even if*, as it seems Wood argues, defiance needs other causal factors to have an effect, i.e. even if the mechanisms in Wood do not work in isolation.

we find that for a given context G^α all conflicts in the set I have the same probability value for mobilization when pleasure of agency is absent, and the same probability value for mobilization when pleasure of agency is present, we can generalize over the set I regarding the relation between mobilization and pleasure of agency.

If the probability of civilians mobilizing in any given particular conflict is the same, given that they feel pleasure of agency and holding all background conditions fixed, then we can construct a big probability space that links pleasure of agency and mobilization independent of reference to any particular conflict. This big probability space then gives us the general causal claim that pleasure of agency causes mobilization. Again, it is important here to note the *ceteris paribus* clause that the background conditions must be fixed; if the conditions are not fixed, the condition that the conditional probabilities are the same will be too strong.

This particular analysis of generalization depends, as argued before, on Hitchcock's particular characterization of the relationship between singular and general causation. There are other questions we may usefully ask about a set of cases I which go beyond the Hitchcock causal homogeneity condition. We might, for instance, otherwise consider what kind of general causal claim we *can* make that applies to the set I, i.e. what kind of similarity condition *is in fact true* of the set I. I will come back to alternative characterizations of similarity in the last section of the paper, after anticipating some complications for Hitchcock's interpretation as applied to the social sciences.

4.2 Evidence for causal homogeneity

So far, I have argued the following. Top-down process tracing, a new method in the social sciences, uses case study evidence in the form of singular causal claims, but also aims at directly inferring general theories, that is, general causal claims. I have emphasized Hitchcock's characterization of the relation between singular and general causal claims, the Hitchcock causal homogeneity condition, which tells us that one must show that the causal relevance of the cause on the effect given a fixed

set of background conditions is the same for all the cases the general theory is meant to cover. In this section, I will discuss how realistic it is to think we may verify this particular characterization of generalization in the social sciences. In particular, I will ask how we can tell whether the causal relevance of the cause on the effect in one case study is the same as the causal relevance of the cause on the effect in another case study. Here, I discuss two complications: the measurement of the conditional probabilities $f_{i\alpha}$ that Hitchcock requires, and the problem of external validity.

4.2.1 Issues with calculating the individual probability spaces

A first empirical complication for the Hitchcock causal homogeneity condition is the difficulty of calculating the individual probability spaces $\{\Omega_i, \mathcal{F}_i, P_i\}$ in any real case. Though discussing how one in practice measures these probabilities is beyond the scope of this chapter, I do wish to point out briefly how my analysis relates to a more formal distinction made in the philosophical literature: constructing an individual probability space can either be done using *conditional* probabilities of type events, or by drawing the probabilities of individual cases from *counterfactual* probabilities of token events (cf. Northcott 2008)[10].

We often use conditional probabilities of type events to construct individual case probabilities when we are working under epistemic limitations; if we do not know the detailed physical history of John's lung, we have to estimate the probability of him developing lung cancer using statistical data about other smokers (such as for instance the data collected in a randomized controlled trial). Appraising probabilities using counterfactuals, on the other hand, means assuming a cause raises the probability of its effect compared to what the probability of the effect

---

[10] This distinction may even apply to the very definition of the probability in question, i.e. a definition in terms of a frequency, always relative to a class, or in terms of a propensity, intended to characterize the single case alone.

would have been in the scenario where the cause is absent. This is not easy; one arguably requires more detailed knowledge of the specifics of John's lung.

When we use conditional probabilities of a type event to shed light on a token event, we are already assuming the general causal claim we wish to corroborate. Thus, a critic may argue, the Hitchcock causal homogeneity condition is not useful in practice, because in order to give evidence that it holds, we must already be assuming it holds implicitly. To avoid this problem, I suggest that in order to give evidence that the causal homogeneity condition is met for some set of individuals, one must calculate the probability space for individuals based on counterfactual probabilities of token events (as indeed fits my analysis in chapter 3), rather than conditional probabilities of the type of event[11].

4.2.2 The external validity problem

Not only is it difficult to establish the individual probability spaces in practice, it is also difficult to give evidence that the homogeneity condition holds. This issue is closely related to the problem of external validity (cf. Guala 2010) which I first mentioned in chapter 1. The problem of external validity is nicely illustrated by the difficulty of extrapolating from a test case to a target case. For instance, in medical research it is difficult to find the conditions under which we can extrapolate the efficacy of a medication in test animals to human patients. In the case of using test animals to comment on human patients, we wish to know that the probability of the effect of the medication is the same, given some set of background conditions, in

[11] I grant that it may still, on occasion, be difficult to construct the probability space for one case without reference to the probability space of other cases. Having constructed the probability space for John and moved on to constructing Mary's space, we may sometimes find new results that make us go back to John's analysis to perform some Bayesian updating on our previous beliefs.

both test animals and humans; part of this difficulty will be establishing the set of background conditions under which the cases will behave similarly.

Analogously, in social science it is difficult to find the conditions under which we can extrapolate the effect of a particular cause in one case study to another, e.g. extrapolate the effect of youth unemployment on the probability of civil war outbreak in Sierra Leone to the effect of youth unemployment on the probability of civil war outbreak in Somalia. This extrapolation requires researchers to assume the causal homogeneity condition outlined above.

We can also frame the problem of external validity by arguing that the *method* we use to test the causal claim that the medication has a particular efficacy in test animals provides no basis for extrapolating the same claim to a different setting, e.g. that of human patients (cf. Cartwright and Efstathiou 2011). In other words, even if the method we have used is apt for establishing singular causal claims, this does not mean it is apt for establishing general causal claims in the form of the causal homogeneity condition.

One might distinguish external validity from generalisation in this way: in the case of external validity, we are driven by our need to know about a particular target case, while in the case of generalisation, we can determine what sort of generalisation we are licenced to make given our evidence base. The set of individuals may or may not include the particular target case that it would be useful to know about. Whereas we can generalize over a wider set of cases I if we find the causal homogeneity condition holds for all $i \in I$, extrapolation from a case i to another case j only requires that a pair i, j is causally homogeneous, without any further reference to a larger set I that contains the elements of this pair. Nevertheless, the problem of external validity has direct consequences for our attempts to give evidence for a generalization over a set I. After all, if our method cannot guarantee the causal homogeneity of even a single pair of elements i and j, how can the method ever guarantee the causal homogeneity condition ranging over a larger set I?

The problem of external validity is complicated by two issues (cf. Steel 2008). The first of these issues is the 'extrapolator's circle', i.e. the idea that we cannot know whether two cases are similar before we have studied them both. In the case of generalization, the 'extrapolator's circle' is indeed a relevant problem because the set I over which we generalize almost always contains more base units of analysis than we have actually studied. The set of all revolutions, for instance, ranges over a wide number of cases, and we do not have detailed information about the mechanisms in all of these cases. The second issue for external validity is that in social science there are always "causally relevant differences between model and target" (Steel 2008, 5). We are never faced with two base units of analysis that have the same background conditions, i.e. we cannot actually observe both $f_{i\alpha}$ and $f_{j\alpha}$ but are more likely to observe $f_{i\alpha}$ and $f_{j\beta}$ (cf. Cartwright and Efstathiou 2011, 232).

There have been some attempts in the literature to solve the problem of external validity (cf. Guala 2010, Steel 2008). The main contender, which derives from philosophy of biology, is a method called 'comparative process tracing' (cf. Guala 2010, Steel 2008, 88-91). (Note that, as stated in the introduction of this thesis, this is a different usage of the term 'process tracing' than the method under scrutiny in chapter 3.) In biology, researchers use comparative process tracing to extrapolate for instance from the workings of a particular medicine in rats to making claims about the workings of that medicine in humans. Comparative process tracing asks the following questions: what stages of the mechanism will most likely differ between the test case (rats) and target case (humans), and at what stages will the cases be similar, based on what we know about other mechanisms (workings of medication) in the test case versus the target case (in rats versus humans)? Are there any focus points in the mechanism that we can use to infer likely differences and similarities at earlier stages? If we can answer such questions, we are one step closer to extrapolation, even if we do not have full knowledge of the target case (cf. Steel 2008, 88-91).

Though these questions may be answerable for biological mechanisms, in social science it is not so clear whether we can access the necessary information. It is unclear whether we can really gain useful knowledge about similarities and

differences between cases, and it is unclear how a fruitful comparison would proceed (cf. Steel 2008, 149-173). This means that the problem of the extrapolator's circle has not yet been solved for the social sciences: we cannot gain knowledge about both $f_{i\alpha}$ and $f_{j\alpha}$. And so, by implication, we cannot gain knowledge about the equality $f_{i\alpha} = f_{j\alpha}$ for *any* i, j in the set *I* we try to generalize over. Until we can solve the problem of external validity, we cannot know whether a potential generalization extends to the target cases of interest.

5. A practical way forward? A second interpretation of generalization

The Hitchcock causal homogeneity condition for generalization can only deal with cases where we have full information about the conditional probabilities for each individual case in the scope of our generalization. The condition does not help us in situations where we have only limited information about these probabilities. However, as the complications above show, it is difficult to collect such information. In the final part of this chapter, I wish to discuss a potential solution to these epistemic problems. After all, despite issues like the extrapolator's circle, social scientists have not given up on the idea of extrapolating or generalizing, and indeed both are not only important for theoretical purposes, but also for, amongst others, policy applications. I discuss Hausman's notion of the average causal effect, a (weaker) alternative to Hitchcock's characterization of generalization that works even in situations where we have limited information. This section provides the backdrop for a deeper analysis of what kinds of general causal claims we *are* often warranted to make in the social sciences, which will take up most of chapter 5.

5.1 Hausman's average effect condition

The Hitchcock causal homogeneity condition is a very strong demand for social science generalization, since it is difficult to find evidence that its requirements are met in any actual cases of interest. Thus, we might find the condition unhelpful in social scientific practice. This is the argument made amongst others by Daniel

Hausman (Hausman 2010) following John Dupré (Dupré 1984). Dupré and Hausman argue that the causal homogeneity condition (which they dub the 'contextual unanimity condition'), is often not what is intended when we make general causal statements, and thus that we need a *different* characterization of the relationship between singular and general causal claims[12].

For instance, Dupré and Hausman show that a general claim such as 'smoking causes lung cancer' will be technically false if we interpret it as saying that the probability of lung cancer increases for *every* individual who smokes. As Hausman points out, there are individuals whose smoking leads to a fatal heart attack before they can even develop lung cancer; these individuals contradict the general causal claim as defined by Hitchcock. Call the set of all individuals whose smoking leads to a fatal heart attack before they develop lung cancer I, and call the set of all other individuals J. Now, define for each $i \in I$ the distribution function $f_{i\alpha}(x) = P_i(E_i|X_i = x, G_i^\alpha)$ as before, i.e. $f_{i\alpha}(x)$ is the probability distribution for individual i developing lung cancer given that i smokes x packs per day under background conditions G_i^α. Define for each $j \in J$ the distribution function $f_{j\alpha}$ in the same way. We will conclude, since all individuals in I die before they develop lung cancer, that $f_{i\alpha} \neq f_{j\alpha}$. Thus, we cannot generalize over the set $I \cup J$ along the lines that Hitchcock presents.

Hausman presents two solutions to this problem. The first is to restrict the general claim's scope; in the example, we may exclude all individuals $i \in I$ who will die of a heart attack before they can develop lung cancer, and make the general causal claim for set J only. However, Hausman points out, this interpretation means we could end up with causal generalizations that "have such a narrow or unclear scope as to

---

be useless" (Hausman 2010, 55). Moreover, "practical causal generalizations (…) typically concern causal relations when the values of other casually relevant variables are *not* unchanging. They concern causal relations in heterogeneous circumstances – that is, across some range of causally homogeneous circumstances." (Hausman 2010, 49)

For that reason, both Hausman and Dupré argue for a second solution: they claim to provide an alternative and more useful characterization of the relation between singular and general causal claims, the *average causal effect*. Using the same notation as before, under this characterization, in population I, $X = a$ as compared to $X = b$ causes E is true (e.g. smoking one pack per day as compared to smoking no packs per day causes lung cancer is true) if and only if two demands are met. Firstly, $P(E|X = a) > P(E|X = b)$. Secondly, the probability difference in the first demand (the fact that for this population, $P(E|X = a) > P(E|X = b)$) must be due to the causal influence of $X = a$ as compared to $X = b$ in some causally homogeneous circumstance occupied by members of the population I. This second demand is meant to exclude cases where the relationship between E and $X = a$ is spurious.

Hausman's alternative is meant to be a *practical* characterization of causal generalizations.

> What is at issue in theorizing about causal generalizations is causal irregularity. The operation of causal factors, whether deterministic or indeterministic, varies from context to context, and guidance is needed when the details concerning the contexts are not known. Theoretical work may focus on individual contexts or homogeneous contexts, because it need not necessarily provide such guidance. But if one hopes to offer advice to people who do not know which homogeneous context they are in, one has to generalize across contexts in which the effects of causal factors are not uniform. (Hausman 2010, 57)

Note that this demand does not ask us about any individual conditional probabilities for the $i \in I$; it simply asks us to compare the average probabilities (which can be calculated using, for instance, observed correlations for population I,

collecting the frequencies with which E occurs in those members of the population for which $X = a$, and for those for which $X = b$). This fits with the motivation for formulating this particular characterization of generalization in the first place: it is useful for those cases where we simply cannot sum over the effects under causally homogeneous background conditions.

Though the average effect condition can be useful in those practical contexts, there are also problems with this characterization of causal generalizations. Hausman accepts that the average effect may sometimes be misleading. If there are many different subpopulations of I, for each of which the relationship between values of X and some effect E differ significantly, then the average effect in I would be misleading. Consider, for instance, a test population in which one type of individual shows strong negative side effects from some medication, and another type of individual does not show any side effects. If we are unaware that there exist two types of person, we may be misled by the average effect to think that the medication only has some moderate negative side effects. Hausman therefore admits that "if the consequences of X for E differ appreciably over different subpopulations, then it can be harmful to generalize over the whole population. It is usually better to generalize concerning the narrowest populations for which the information is available." (Hausman 2010, 59)

Let me sum up. Hausman argues against Hitchcock's interpretation of causal generalizations, the causal homogeneity condition, because this condition requires one to know how to partition a background into causally homogeneous classes, which requires us to know already what other factors influence the effect of interest E besides the cause X. However, Hausman also warns us that his more practical interpretation of generalizations, the average effect condition, can be misleading if the heterogeneous backgrounds are *too* different. While it is difficult to establish a Hitchcockian generalization in the social sciences, these generalizations are useful in decision making; it is much easier to establish a Hausmanian generalization in the social sciences, but these generalizations are less useful in decision making.

6. Conclusion

In this chapter I have analysed the relationship between singular case study evidence and general hypotheses in the social sciences, firstly by arguing that generalization is indeed an aim of some social scientists, and secondly by asking what evidence one needs to substantiate such general hypotheses. I looked, in particular, at one interpretation of 'general', derived from Hitchcock's characterization of the relationship between singular and general causal claims. I outlined that this characterization requires us to prove, for all elements of the set we wish to generalize over, that the conditional probabilities of the effect given the cause, all other things being equal, are higher than the conditional probabilities of the effect given the absence of the cause. I formalized this in terms of a little probability space and a big probability space, following Hitchcock.

I then discussed to what extent Hitchcock's interpretation is realistic for the social sciences. I focused on several causal complications, including the external validity problem and the extrapolator's circle problem, all of which mean that it is difficult to establish the exact little probability space for each individual case of interest, and thus to establish the more general big probability space for the entire set of cases of interest.

In light of these complications, I briefly discussed one alternative interpretation of the relationship between singular and general causation, due to Hausman and Dupré, i.e. the average effect condition. This condition is supposed to be applied in cases where we do not have any insight into the distinct causally homogeneous subpopulations in the set of cases of interest. Though practically more feasible, because this condition explicitly averages out over potentially causally very different subpopulations, I argued it may be misleading.

What, then, are we left with? There are two aspects to evaluating whether one ought to accept Hitchcock's or Hausman's interpretation of generalization. The first aspect is descriptive: we may ask, for any generalization made in the social sciences, whether it meets the requirements of Hitchcock's causal homogeneity condition (whether for every case in the sample set, all other things being equal, the

conditional probabilities of the effect given the cause are the same) and whether it meets the requirements of the Hausman average effect condition (whether on average, the probability of the effect given the cause is larger than the probability of the effect in the absence of the cause, and that this is due to the relation between the cause and effect for at least a substantial subpopulation). This is an empirical question that we can ascertain for each generalization individually, and as such I want to set it aside here, though I will emphasize again that due to such complications as the external validity problem and the extrapolator's circle, for many cases it would seem the Hausman average effect condition is more easily determined in practice.

The second aspect on which we can evaluate Hitchcock versus Hausman is normative: whether, in the social sciences, we *should* attempt to find evidence for the stronger Hitchcock causal homogeneity condition, or just for the weaker average effect condition. In order to answer that question, I will set up a synthesis between the two accounts in the next chapter, chapter 5. There, I argue there that one way of weakening the Hitchcock causal homogeneity requirement is to look for *epistemic* homogeneity rather than ontological homogeneity: we may assume homogeneity over a set until proven otherwise.

Chapter 5

Epistemic Homogeneity: a Taxonomy

1. Introduction

In the previous chapter, I spelled out two different interpretations of generalization: the Hitchcock causal homogeneity condition, which requires each unit generalized over to have the same probability of the effect given the cause, all other things being equal, and the Hausman average effect condition, which allows us to average over causally distinct subpopulations.

As I argued in the last chapter, Hausman's average effect condition gives population-level information. The condition will be less useful for deriving information about individual cases, e.g. if there are significant subpopulations of P in which the causal relationships between values of the putative cause X and effect of interest E differ widely. I hinted at Hausman's claim that in those cases, "[i]t is usually better to generalize concerning the narrowest populations for which the information is available" (Hausman 2010, 59).

In this chapter, I will provide a synthesis of the two interpretations of generalization. As I have argued in the previous chapter, the two interpretations are more pertinent to different epistemic situations: in the one, we are aware of all the relevant probabilities for individual cases in our population of interest and so we have evidence for the stronger generalization, and in the other, we are unaware of the specific probabilities and so we merely have evidence for a weaker generalization. Synthesizing the two, I will argue for pursuing *epistemic* homogeneity rather than ontological homogeneity, i.e. one should assume homogeneity until proven otherwise.

In this chapter, I will first spell out this epistemic homogeneity condition in more detail, following an argument that dates back to Salmon (1971), and formalize it in terms of the framework by Hitchcock that I introduced in the previous chapter. I

then illustrate this with an example from the civil war literature, the research into the homogeneity of the class of all civil wars by Nicholas Sambanis (2001). Second, I refine the notion of epistemic homogeneity, by distinguishing between epistemic homogeneity relative to *one particular property*, and total epistemic homogeneity relative to *all relevant properties*, i.e. a complete causal picture. Third, I argue that what one considers to be an epistemically homogeneous class depends not only on the causal structure of the elements of that class, but also on pragmatic considerations of the resources available to treat individuals as different. Thus, I again taxonomize the epistemic homogeneity condition, this time into *true* epistemic homogeneity, and *pragmatic* epistemic homogeneity.

2. Epistemic homogeneity defined

2.1. Wesley Salmon's reference class rule

In this section, I first describe Wesley Salmon's notion of epistemically homogeneous reference classes for probabilistic explanation, and then use this as a basis to formalize the epistemic homogeneity condition for *causal* explanation using Hitchcock's framework from the previous chapter.

In his 1971 paper "Statistical Explanation and Statistical Relevance", Wesley Salmon proposed the distinction between objectively and epistemically homogeneous reference classes in an attempt to solve the reference class problem for probabilistic explanation. The reference class problem is the issue that calculating a probability of an event requires us to specify a reference class for that event, whilst there is no straightforward 'correct' reference class but rather a variety, each of which gives a different probability. For example, to calculate the probability that John will die of a heart attack, we must know what reference class to use: the class of all Caucasian men? All middle-class people? All professional tennis players?

Salmon solved the reference class problem by arguing for the use of the 'broadest homogeneous reference class'. Let me first introduce the concept of a homogeneous reference class with an example. Assume we are investigating a class of events or

phenomena, A, and are interested in the frequency with which some property B occurs in that class. For instance, assume we are investigating the class of all English sixth form pupils, and are interested in the frequency of A* results amongst them. Class A is called a 'homogeneous reference class' for property B if there is not a property C according to which we can partition the class further such that the probability of having property B is different in the resulting subclasses (i.e. if it is not the case that $P(B|C, A) \neq P(B|A)$). (Any partition that refers to property B will always be a partition that results in a different frequency of B; that is why we exclude such partitions.) So, for instance, if there is no way of dividing up the class of all English sixth form pupils, so that in one subclass the frequency of A* results is significantly different than in the other subclass, then we call the class of all English sixth form pupils *homogeneous with respect to getting A* results*. Now, of course the class of all English sixth form pupils is *not* homogeneous with respect to getting A* results; we know, for instance, that students from underprivileged backgrounds typically do worse than students from other backgrounds, i.e. we know that $P(\text{A*results}|\text{underprivileged background, English sixth form pupil}) <$
$P(\text{A*results}|\text{English sixth form student})$.

Consider a few more examples. The class of all pregnancies is not homogenous for the property of babies born with physical, developmental, and functional problems, because we can further divide the class according to the place selection that picks out mothers who drink more than two glasses of wine per day. The class of university graduates is not homogeneous for the property of income, because we can further divide the class according to the place selection of gender.

A first thing to note is that, at least in the university graduates example, we could also partition the class with other place selections, e.g. in this case according to social class. It is not always clear which of these classes we should work with, which is why Salmon argues for investigating "the broadest homogeneous reference class to which the single event belongs" (Salmon 1971, 43), i.e. the homogeneous class

with the most members[1]. For Noa, a female graduate from a working class background, this homogeneous class will be the class of all female working-class graduates. For Clarence, a male graduate from an upper class background, this homogeneous class will be the class of all male, upper-class graduates. Note that if it were to later turn out that the highest level of education attained by Noa and Clarence's parents also matters to their income, this property would have to be included in the delineation of their homogeneous class, and thus we would have to narrow the homogeneous classes further.

A second thing I wish to stress is that in Salmon's framework, a class A is never simply homogeneous – it is homogeneous with reference to a particular property under investigation. This means that a class that is homogeneous with reference to one particular property B may not be homogeneous with reference to another property C. Thus, whereas a theory concerned with property B may have scope conditions a_1 (may only apply to entities with property a_1), a theory concerned with property C may have scope conditions a_2 (may only apply to entities with property a_2). For instance, if true, the democratic peace theory (the theory that democracies are hesitant to go to war with each other) has all (pairs of) democracies as its scope, since it claims this class of democracies is homogeneous with respect to the probability of war onset. For other theories, however, we will be forced to partition the class of democracies and thus give the theory a more narrow scope; for instance,

---

[1] Readers might compare this with Hans Reichenbach's suggestion to consider the *narrowest* reference class; Salmon's 'reference class rule' is a reformulation of Reichenbach's method. As Salmon writes, "The aim in selecting a reference class to which to assign a single case is not to select the narrowest, but the widest, available class. However, the reference class should be homogeneous, and achieving homogeneity requires making the reference class narrower if it was not already homogeneous. I would reformulate Reichenbach's method of selection of a reference class as follows: choose the broadest homogeneous reference class to which the single event belongs." (Salmon 1971, 43)

a political scientist interested in the causes of party polarization may wish to distinguish multi-party democracies from two-party democracies.

The idea that a class is homogeneous with reference to a particular property is nothing new (cf. Clarke 2011). For instance, this idea underlies the following example from medical diagnosis[2]. There exists a variety of different rheumatoid ailments, many of which can be treated with prednisone. When it comes to finding a treatment for any individual patient, it is sufficient to know whether their ailment is belongs to the group susceptible to improvement due to prednisone. We do not need to know which particular form of rheumatoid ailment they have. The broadest homogeneous reference class relative to the property of being susceptible to improvement due to prednisone will be the class of prednisone-susceptible rheumatoid ailments.

However, in the social sciences as in medicine, researchers are often unaware of the full causal picture, i.e. which properties are relevant for the cause and effect under discussion, and which are not. Salmon's suggestion, in those cases of 'incomplete information', is as follows: "[w]hen we know or suspect that a reference class is not homogeneous, but we do not know how to make any statistically relevant partition, we may say that the reference class is epistemically homogeneous" (Salmon 1971, 44).[3]

How does this solution to a problem for probabilistic explanation relate to the interpretations of generalization I have outlined in the previous chapter? There, I discussed Hitchcock's (probabilistic) interpretation of how one generalizes from singular causal claims to general causal claims, and applied this theory of

---

[2] Thanks to Luc Bovens for pointing me to this simple example.

[3] We may also study classes that are epistemically homogeneous, rather than ontologically homogeneous, for pragmatic reasons: we know that many entities in social science are not homogeneous, but we may treat them as if they were for some purposes. I will discuss this pragmatic consideration in the last section of this chapter.

generalization to the social sciences before concluding that it is difficult to find evidence for this causal homogeneity condition in social science. I then outlined Hausman's demand for using the average causal effect as long as subpopulations do not diverge too much. These two interpretations, I will now show, are at bottom complementary: they are meant to apply to different epistemic situations. In the one, we are aware of all the relevant probabilities for individual cases in our population of interest, and in the other, we are unaware of the specific probabilities. If we take Hausman's interpretation as a suggestion to look for *epistemic* homogeneity rather than ontological homogeneity, i.e. to assume homogeneity until proven otherwise, we will be able to make epistemic progress regarding a set of cases.

I will now express epistemic homogeneity formally using last chapter's framework. Then, I will turn to a refinement of these ideas by distinguishing complete epistemic homogeneity from epistemic homogeneity relative to a particular property. Last, I will present a second dimension to the taxonomy of epistemic homogeneity, the distinction between epistemic homogeneity for practical purposes (when we do not have the resources to distinguish between individuals) and true epistemic homogeneity.

2.2 Epistemic causal homogeneity in a formal framework

Let me return to the framework from the previous chapter, in order to clarify what epistemic homogeneity would amount to there. Recall that for Hitchcock's causal homogeneity condition, we need the following ingredients. Hitchcock constructs a 'little probability space' $\{\Omega_i, \mathcal{F}_i, P_i\}$ for every individual i in the set we wish to generalize over, consisting of a sample space Ω_i which contains all possible instances $X_i = x$ of the process being modelled (for example, as we have seen, smoking x packs per day), a set $\mathcal{F}_i$ which contains 'events', i.e. combinations of outcomes from Ω_i (for example, as we have seen, smoking more than x packs per day), and a probability function $P_i: \mathcal{F}_i \rightarrow [0,1]$ to model the conditional probability of the effect E_i of interest (e.g. the probability of contracting lung cancer) given an

event from $\mathcal{F}_i$. Moreover, Ω_i can be partitioned into background contexts $G_i^1, G_i^2, \dots$. The 'little probability space', then, is a distribution function $f_{i\alpha}(x) = P_i(E_i | X_i = x, G_i^\alpha)$ that tells us the probabilities of the effect E_i given that the causal variable X_i has value x and given background context G_i^α. A general causal claim, under Hitchcock's interpretation, can be made over the set of those individuals for which the little probability space is the same, i.e. for those sets I of which all elements i, j have the same distribution, i.e. $f_{i\alpha} = f_{j\alpha}$.

In cases when not all elements i, j of the set I we are interested in have the same distribution function, the scope of the population I we generalize over must be restricted. In the simplest case, there is one particular property C which has led to the problem. Now, following Salmon's terminology, it is this property C that we could use to partition I into two homogeneous classes. The result will be two subsets of I, $I_1 \subset I$ and $I_2 \subset I$ (with, seeing as this is a partition, $I_2 = I \backslash I_1$, i.e. $I_2 = \{x \in I | x \notin I_1\}$), which each in turn are causally homogeneous but for which it is the case that for every $i_1 \in I_1$ and for every $i_2 \in I_2$, $f_{i_1} \neq f_{i_2}$.

So far this framework cannot deal with cases of incomplete information. This is a problem because researchers are hardly ever in possession of all relevant information about the distribution functions $f_{i\alpha}$. Researchers do not know which properties of an individual $i \in I$ might disrupt the causal relation between X_i and E_i, nor which properties might support it. We may, in those cases, not wish to bet that the strict generalization is true; we may prefer to instead adopt a weaker notion of generalization of which we are relatively sure it is true in cases of interest.

So, in general, if researchers have incomplete information about the causal structure of the area they are investigating, they might not be able to come up with a partitioning C. It is exactly this 'incompleteness' that Salmon captures with his notion of epistemic homogeneity. As long as they do not have a partitioning C that they know will show that the class I is heterogeneous with respect to the causal relation of interest, researchers ought to call that I epistemically homogeneous. For that set I we can use the Hausman average effect condition to make general claims, since it does not require further information about the individuals in that set; once

we find more information, we can then (following Hitchcock) narrow down the set further. This synthesis between the two frameworks for generalization, I believe, is key to making epistemic progress in causal inquiry in the social sciences.

3. Example: civil war studies' search for epistemic homogeneity

In this section, I wish to discuss an example of the search for causal homogeneity, viz. the move from a general theory on civil war onset to a more specific theory on ethnic civil war onset by Nicholas Sambanis. I will show that the class of states at civil war was an epistemically homogeneous class with respect to several properties, including economic and political factors, until Nicholas Sambanis figured out a way to make a statistically relevant partition in the class, namely between ethnic and non-ethnic civil wars[4]. I will make clear what the results of the partitioning of this class were, and thereby illustrate the notions of epistemic homogeneity and epistemic progress discussed above.

3.1 Civil wars as an epistemically homogeneous class

In 2001, Nicholas Sambanis asked whether ethnic and non-ethnic civil wars start for the same reasons. The ordinary theories in the civil war literature at that time assumed so: the then prominent 'economic' theory of Collier and Hoeffler (Collier and Hoeffler 2000), which I have discussed in chapter 2, aggregated civil wars since

---

[4] Note the following difference between the study of civil war and the study of cancer patients as in the above: in the case of the causal link between lung cancer and smoking, we examined the *effect* properties of the class, i.e. we examined which persons in the class contracted lung cancer; in the civil war case, we examine the *causal* properties of the class, i.e. we examine which states in the class have the same causes.

1960 into one class[5] and suggested as I have shown that such wars start mainly because of economic factors (such as financial incentives for the rebels) rather than political factors (such as the level of democracy of the country and of its neighbouring countries).

In the terminology introduced in the previous section, one may say that Collier and Hoeffler treated the class of all civil war onsets as an epistemically homogeneous reference class with respect to the properties under investigation (i.e. with respect to the potential causes). Sambanis' contribution to the civil war literature was suggesting that there might exist a relevant partition, 'ethnicity', which divides the class of civil war onsets into causally dissimilar subgroups. In this section I will highlight the differences between ethnic and non-ethnic civil wars in Sambanis' framework. I briefly outline the aggregate theory of civil wars as presented by Collier and Hoeffler (Collier and Hoeffler 2004), and then discuss the properties relative to which Sambanis believes there is a partitioning which shows the ethnic heterogeneity of the class of all civil wars.

Collier and Hoeffler and Sambanis define 'civil war' in a similar manner, based on the definition used by the Correlates of War (COW) database outlined in chapter 2. As indicated, Collier and Hoeffler do not distinguish between different types of civil war, but instead consider the class of civil wars as causally homogeneous. They discuss this decision in later writing, noting that although some data sets define conflicts in terms of the underlying issues (as is the case when scholars make a distinction between ethnic and non-ethnic civil wars) they have decided not to do so because "the classification of conflicts according to their causes does not seem helpful (…) if we want to analyse the causes of civil war" (Collier and Hoeffler 2001, 5).

---

[5] As stated in chapter 2 (footnote 1), I will not discuss this temporal element to theories on the causes of civil war here; the choice for a data set that starts in 1960 may indicate that scholars had sufficient reason to treat the wars before the 1960s as having a significantly different causal structure.

In terms of the interpretation outlined in the previous section, Collier and Hoeffler consider the class of all civil wars, *A*, and search for the frequency with which property *B*, 'ethnic fragmentation' (a potential causal variable), occurs in that class (which is then compared with the frequency with which property *B* occurs in a similar class where civil war did not break out). According to Collier and Hoeffler ethnic fragmentation is a measurable proxy for the 'coordination costs' of a rebellion. Higher coordination costs, they argue, are an inhibiting cause of civil war onset. Ethnic fragmentation is also related to civil war onset via a different route: according to Collier and Hoeffler, ethnic dominance leads to minority oppression, which in turn increases the risk of civil war onset.

3.2 Partitioning civil wars into ethnic and non-ethnic wars relative to ethnic heterogeneity

In contrast, in "Do Ethnic and Non-ethnic Civil Wars Have the Same Causes?" Nicholas Sambanis argues for partitioning *A*, the class of civil wars, in relation to property *B*, 'ethnic fragmentation'[6], according to the place selection *C* of 'ethnicity'[7].

---

[6] Sambanis uses a similar but slightly more inclusive notion of ethnic fragmentation than Collier and Hoeffler do, which combines racial, linguistic, and religious considerations. This notion is called ethnic heterogeneity (EHET) in short and follows Vanhanen (1999). EHET ranges from 0 (extreme homogeneity) to 177 (extreme heterogeneity). Sambanis argues for using EHET rather than ethno-linguistic fractionalization (ELF) because EHET data is both more recent and available for more countries. Besides ethnic fragmentation, Sambanis also investigates the correlation between several other (proxy) variables and (ethnic) civil war onset, thus considering different properties *B* for the same class *A*. The different *B*s include democracy, war occurrence in the country's region, and the end of the Cold War. I will not discuss these here.

Having thus partitioned the class of civil wars into ethnic and non-ethnic civil wars (the definition of which I will outline below), Sambanis uses statistical methods to test the hypothesis that the probability of ethnic war breaking out is an increasing function of the degree of ethnic fragmentation[8]. He finds that ethnic fragmentation is not significant in any of the regressions which treat civil war as a homogeneous class, including the regressions in Collier and Hoeffler's study. After partitioning, however, the numbers look significantly different. Ethnic fragmentation "is among the most robustly significant variables and is positively correlated with the onset of ethnic war: as a country becomes more heterogeneous, the probability of occurrence of an ethnic war increases" (Sambanis 2001, 273). However, for non-ethnic war onset, ethnic fragmentation is no longer statistically significant.

Sambanis defines ethnicity following earlier work by Donald Horowitz (1985), in which an ethnic group can be defined in terms of anything from "colour, appearance, language, religion, some other indicator of common origin, or some combination thereof" (Horowitz 1985, 17-18) and covers other terms like "tribes, races, nationalities, and castes" (Horowitz 1985, 53). A civil war is an *ethnic* civil war, Sambanis argues, if the core issues in the conflict are "integral to the concept of ethnicity" (Sambanis 2001, 261-262). Or, alternatively, an ethnic war is a "war among communities (ethnicities) that are in conflict over the power relationship that

---

[7] Sambanis' motivation for investigating ethnicity includes evidence from studies of civil war termination (as opposed to onset), which suggested there are substantive differences between ethnic and non-ethnic wars, as well as evidence from the literature on ethnicity and ethnic conflict in general. Although experts on ethnicity and ethnic war all "share a conviction that ethnicity is a critical variable in explaining civil violence" (Sambanis 2001, 263), Sambanis writes that no one had attempted to prove suspicions about there being substantive differences between (the causes of) ethnic and non-ethnic war onset as well.

[8] Note that this is not quite the same as finding the frequency of ethnic fragmentation amongst the class of ethnic civil wars.

exists between those communities and the state" (Sambanis 2001, 261). He codes a conflict as an ethnic civil war if it is a civil war (for exact requirements, see (Sambanis 2001, 262)) and if it counts as an episode "of violent conflict between governments and national, ethnic, religious, or other communal minorities (ethnic challengers) in which the challengers seek major changes in their status" (Sambanis 2001, 262).

In light of the discussion of Salmon's reference class rule, whether one can call 'ethnicity' a proper partitioning for the class of civil wars in relation to the property of ethnic fragmentation is dubious. At first glance, one may suspect that taking 'ethnicity' as a place selection breaks the rule that a partition of class A is only a place selection if it does not refer to the property B under investigation; or, to repeat Collier and Hoeffler's words, we cannot classify conflicts "according to their causes (…) if we want to analyse the causes of civil war" (Collier and Hoeffler 2001, 5).

On the other hand, ethnic fragmentation refers to the number of different ethnic groups within a state. Thus, it is not the case that the partition into 'ethnic' and 'non-ethnic' civil wars refers to the property of 'ethnic fragmentation'. As Sambanis puts it, "not all wars that involve ethnic groups as combatants should be classified as ethnic wars. The issues at the core of the conflict must be integral to the concept of ethnicity" (Sambanis 2001, 261-262). A country can in theory have a low degree of ethnic fragmentation and still be at ethnic civil war; two ethnicities is all it takes. And, vice versa, a country can have a high degree of ethnic fragmentation without an ethnic civil war being fought there – and if there is a civil war being fought in the country, it may have started for different reasons[9]. For those reasons, one might argue that 'ethnicity' is a proper partitioning, despite first appearances to the contrary.

---

[9] This idea is proposed by others in the political science literature as well; Wimmer, Cederman, and Min (2009) for instance argue that "highly diverse societies are not more conflict prone", but that it is particular ethno-political balances of power in a society that make it more conflict prone.

However, the correlation between ethnic heterogeneity and ethnic civil war is arguably not unexpected given that both are defined in terms of ethnicity; they might not be interchangeable but they are closely related. Unless we can show the conceptual independence between the two it is not useful to make a distinction between ethnic and non-ethnic civil wars merely on the basis that they have a different causal connection to the property ethnic fragmentation. What Sambanis needs to show, I would argue, is that ethnic wars are different from non-ethnic wars in a way that goes beyond their causal history of ethnic fragmentation. I will discuss this move in the next section.

3.3 Partitioning into ethnic and non-ethnic civil wars relative to other causes

Though it is not immediately obvious that Sambanis' distinction between ethnic and non-ethnic wars is a proper partitioning for the property of 'ethnic fragmentation', there are other properties that Sambanis investigates. He shows that ethnic and non-ethnic civil wars also differ in relation to those other properties, i.e. that the class of 'civil wars' is heterogeneous relative to other properties besides the (dubious) ethnic heterogeneity.

Sambanis considers several such properties, including political variables such as the polity score of the country, the polity score of its neighbouring countries, and economic variables such as real per capita income. Sambanis shows that whilst the polity score of a country is statistically significant for ethnic war onset[10], its polity is

---

[10] One may suggest that it is unsurprising that the polity score of a country influences the probability of ethnic civil war onset, speculating it would be more challenging to get a high polity score if the country's ethnic heterogeneity (EHET) score is high. However, Sambanis investigates this as part of his statistical work, showing that the (negative) correlation between ethnic heterogeneity and polity is very weak (Sambanis 2001, 271, table 2).

non-significant to non-ethnic war onset (Sambanis 2001, 276). Moreover, real per capita income is more significant to non-ethnic war onset than it is to ethnic war onset, leading Sambanis to speculate that economic variables are a more important causal factor in non-ethnic war onset than in ethnic war onset. He summarizes his conclusion as follows:

> Overall, there are some important differences between ethnic war onset and the onset of civil war more generally (i.e. aggregated civil wars and/or revolutionary wars specifically). Politics is more important than economics in causing ethnic civil war, and ethnic heterogeneity significantly increases the risk of such war. (Sambanis 2001, 279)

In other words, "[ethnic] wars are predominantly caused by political grievance, and they are unlikely to occur in politically free (i.e. democratic) societies" (Sambanis 2001, 280).

3.4 Lessons taken forward

So, I have shown in this section that civil war onset was treated as a homogeneous class, despite indications to the contrary from other areas of the civil war literature. Sambanis showed that for the property of ethnic fragmentation, the class could be fruitfully partitioned into two causally dissimilar subgroups, i.e. ethnic and non-ethnic civil wars. I have analysed to what extent this partitioning is a proper partitioning. I have also shown that Sambanis' partitioning into ethnic and non-ethnic civil wars was relevant for *other* properties which are more straightforwardly independent of ethnicity, i.e. political and economic factors. If we accept Sambanis' statistical results, we must also accept that general theories of the causes of civil war cannot have all civil wars as their scope; the scope conditions had to be limited to ethnic civil wars or non-ethnic civil wars (not both).

It is worth noting that since Sambanis' article came out, there have been other attempted taxonomies of civil war, including further partitioning of ethnic and non-ethnic civil wars (cf. Horowitz and Heo 2008). Further research into such partitions may lead to smaller epistemically homogeneous sets, e.g. of "ethnic internal

conflicts (…) fought over issues such as ethnic self-determination and control in territories defined as homelands, or over distributing power and choosing policies based on collective goals of ethnic groups" (Horowitz and Heo 2008, 2).

This case study of Sambanis can be taken to reiterate that we must check for the independence of the place selection and property, a requirement I introduced in my analysis of Salmon's reference class rule in section 2.1. But the case study also highlights a relevant distinction in the notion of epistemic homogeneity: whether we consider epistemic homogeneity relative to one particular property, or *total* epistemic homogeneity. In the former case, researchers are merely interested in a particular cause, and in the latter, researchers aim to find a complete causal picture of a particular social phenomenon. I will discuss this distinction in more detail below.

4. Further refinements of the notion of epistemic homogeneity

4.1 Classes that are epistemically homogeneous relative to more than one variable

So far, I have argued that if we do not know what the right partitions in a class are to show causally relevant differences, then we may call this class epistemically homogeneous. We might say that things are equal until proven different[11]. Without such a ruling, any policy that requires some assumption about the probabilities of all cases under its scope will be incomplete. As already indicated when I discussed Hausman's average effect condition, assuming a population is homogeneous will have more serious consequences if there turn out to be large causally relevant

---

[11] This rule is therefore similar to the principle of insufficient reason, in the sense that in both cases, we declare that if there is no knowledge indicating unequal probabilities (causal relevance) for entities in our set, we will not distinguish between these entities.

differences in subpopulations; if, for instance, territorial ethnic civil wars respond quite differently to international interventions than non-territorial ethnic civil wars do, then the average effect of international interventions for the class of all ethnic civil wars will be a misleading source of information for anyone deciding whether or not to intervene in an individual conflict.

In the case study above, I found that there are cases when a researcher does not simply investigate the homogeneity of a population relative to one particular variable. As I showed also in chapter 2, there are instances when researchers are not simply interested in the relationship between one property and the class under consideration (as when we try to investigate whether the polity score of a country is a (contributing) cause of civil war). Instead, researchers may wish to link a whole list of properties to the class under consideration (as when we are interested in a complete causal picture of civil war). Sambanis, in his case study, looks at more than just ethnic fragmentation; he concludes that there are "statistically significant differences between the means of core variables (e.g. political rights, ethnic heterogeneity, and war duration) sorted by war type" (Sambanis 2001, 272). This leads him to argue for researching the differences between ethnic and non-ethnic wars in more detail. We may call the sort of homogeneity we are looking for when it comes to a whole list of properties 'total causal homogeneity', as opposed to the 'relative causal homogeneity' researchers are looking for when they only consider one (potentially causal) property.

A critic might respond that we have already come across the argument, in chapter 2, that there are no 'social kinds', i.e. we have discussed reasons to believe that all social science cases have causally relevant differences (cf. Little 1991). This means that total ontological causal homogeneity is far too strong a requirement for social science. Nevertheless, I would argue, one can fruitfully employ a notion of total *epistemic* causal homogeneity. In other words, though after comparative research it will always turn out that all homogeneous classes have size one, i.e. contain only one element, researchers may work with what they believe to be a completely causally homogeneous class until they find a relevant partition.

Consider the following formalization of this search for total causal homogeneity. Looking for total *ontological* causal homogeneity would involve the following. Call $X_i = \{X_i^1, X_i^2, \ldots, X_i^n\}$ a set of variables about which we wonder whether they are causally related to an effect of interest E_i for each individual i in a population I, the further properties of which we describe by a background context G_i^α (defined before as a subset of Ω_i). Now index the little probability space for any $i \in I$ not only to the indivuidal i, but also to the causal variable X_i^k of interest, as such: $f_i^k : \{x\}_k \rightarrow P_i$, with $f_i^k = P_i(E_i|X_i^k = x, G_i^\alpha)$. A set I of individuals will then have total ontological causal homogeneity if for any $i, j \in I$, for all X_i^k, X_j^k, it is the case that $f_i^k = f_j^k$. If this total ontological causal homogeneity condition is not met, in the simplest case we will be able to find a partition C of the population I into two subpopulations, I_1 and I_2, such that for all $i_1, j_1 \in I_1$, for all $X_{i_1}^k, X_{j_1}^k$, it is the case that $f_{i_1}^k = f_{j_1}^k$, but this is not the case for any $i_2 \in I_2$. That is, $f_{i_2} \neq f_{i_1}$.

Total epistemic causal homogeneity, then, is the weaker notion of generalization which we adopt when we do not have sufficient information to formulate such partitions C. In doing so, however, we must respect that any stronger notion of generalization will be inapplicable in the social sciences, since as stated above, all homogeneous classes in the social sciences have size one.

4.2 Pragmatic epistemic homogeneity

The criticism that there always exist causally relevant differences between social phenomena now brings me to another way in which we can taxonomize epistemic causal homogeneity. Though one may agree that there are always going to be causally relevant differences between cases, this has so far not stopped social scientists from attempting to generalize, whether relative to one particular property (in the case of epistemic causal homogeneity relative to a property) or relative to all properties (in the case of complete epistemic causal homogeneity). In this section I will highlight what reasons researchers can have to generalize despite these relevant differences, and distinguish between 'true epistemic causal homogeneity' (when researchers truly cannot think of a relevant partitioning of a class) and

'pragmatic epistemic causal homogeneity'. Researchers are working with the latter when they know particular partitionings of the class exist, but treat the class as homogeneous anyway for pragmatic reasons (including, for instance, not having the time, energy, or resources to adjust policy to any subclasses). In the latter case, the causal claims one makes about a class are not tracking the true causal structure of the world, or even what we believe may be the true causal structure given epistemic limitations, but such claims are nonetheless useful simplifications[12].

In order to highlight these two constraints on homogeneity considerations, consider the following example. If we wish to know whether there is a type-level causal relation between students' ethnicity and their educational attainment, our conclusions in chapter 2 tell us to first decide on the systematization of these concepts. (We might find e.g. maths and English skills most important, subsequently systematizing educational attainment by requiring all students to pass two standardized tests, the GCSE in maths and English. We might define ethnicity by surveying the individual, asking them to choose one or more ethnicities from a list, e.g. 'white Irish'). The second step to finding a type-level causal relation between 'ethnicity' and 'educational attainment' will be to measure the relation between the two variables for a particular population: e.g. we may calculate the correlation between ethnicity and educational attainment for all Key Stage 4 pupils in England.

Now, crucially, assuming that each ethnic group (as defined by the survey) in English Key Stage 4 pupils is causally homogeneous with respect to educational attainment will rely on two considerations. Firstly, as I have argued in the first part of this chapter, such an assumption will depend on whether researchers have any evidence to the contrary (whether there exists a place selection that would show there exist causally heterogeneous subgroups within at least one ethnic group of

---

[12] This division between pragmatic and true epistemic homogeneity is another example of the interplay between *causal relations* and *human aims and values*, which was the topic of chapter 2.

English Key Stage 4 pupils with relation to the property of educational attainment; think of e.g. gender).

Secondly, which group researchers treat as causally homogeneous will depend on the purpose for which they seek the causally homogeneous group. To name but a few aim-related considerations, which group researchers treat as causally homogeneous will depend on whether they can afford (the time, energy, finances) or risk to generalize over causally heterogeneous subgroups. Do they have the time, energy, or finances, say, to research educational attainment for each ethnic subgroup? What would they risk if they did not, and what might the consequences be? Moreover, even if they have good reasons to believe there exist causally relevant differences between some subgroup (differences, for instance, between the educational attainment of Ciara, a girl, and Liam, a boy, both white Irish pupils in Key Stage 4), do the researchers have the time, energy, or finances to treat each subgroup differently (to give Ciara and Liam different learning activities or preparatory tests, for instance)?

The distinction between true and pragmatic causal homogeneity is also related to the following distinction, which was first introduced in chapter 4. Generalization always amounts to describing properties of some population. On some occasions, all one cares about are facts about the population at large; in that case, averaging out over subpopulations we know to be heterogeneous is acceptable, i.e. we can work under pragmatic epistemic homogeneity. On other occasions, one cares about painting as accurate a picture as possible for each individual in the population, and not just for the population at large. In that case, averaging out is unacceptable; we will not work under pragmatic epistemic homogeneity.

The distinction between true and pragmatic epistemic homogeneity raises a few questions. Firstly, we may ask whether true epistemic causal homogeneity can still be called *epistemic* (as opposed to ontological). I would argue that it can; it is a description of the state of our knowledge about the causal structure of a particular class of units, and not a description of the state of the world. Secondly, we may ask whether there is any reason to use the term 'true epistemic causal homogeneity'

given that, as I have admitted on several occasions in chapter 2, there are no social kinds: there are always causally relevant differences between units researchers aim to generalize over. I would argue there are reasons to use the term: it describes the state of our current knowledge. To make a distinction between *true* epistemic causal homogeneity and *pragmatic* epistemic causal homogeneity allows us to highlight the difference between *not knowing* a way to subdivide a group into idiosyncratic individuals, and *knowing* such a subdivision exists, yet not *acting* upon it.

5. Conclusion

In the above, I have argued that ontological causal homogeneity is too strong a requirement for building general theories. It requires one to have knowledge about all the events or phenomena in one's class of interest, which does not reflect scientific practice. I defended one particular way of weakening the notion of homogeneity so as to better understand the process of generalizations in social science, i.e. making a distinction between ontological and epistemic homogeneity. I showed how the notion of epistemic homogeneity can bring us closer to accounting for what constitutes an adequate justification for the scope conditions of a general theory, and illustrated this with the development of Nicholas Sambanis' theory on ethnic civil wars. I then refined the notion of epistemic homogeneity by showing one can delineate a class of events or phenomena according to one property (is this class of pupils homogeneous with reference to their educational attainment?) or according to all properties (is there class of civil wars homogeneous with reference to all causally relevant variables?), i.e. I distinguished between relative epistemic homogeneity and total epistemic homogeneity. Moreover, I discussed and illustrated a further distinction between pragmatic epistemic homogeneity, which takes into account the practical constraints on how much time and effort one can put into distinguishing a group into subgroups, and true epistemic causal homogeneity, which merely refers to the case when a researcher does not know a relevant partitioning.

Chapter 6

Thinking about Social Mechanisms

1. Introduction

In this chapter, I analyse to what extent the prominent 'new mechanist' theory of causal mechanisms developed for philosophy of biology by Peter Machamer, Lindley Darden, and Carl Craver (MDC)[1] applies to mechanistic reasoning in the social sciences. One of the central arguments in the original MDC is that biological mechanisms explain in virtue of the fact that the mechanisms' components, i.e. 'entities' and 'activities', are organized in such a way that they are productive of regular changes. In this chapter, I will argue that though there are many parallels in the way mechanisms are described in Machamer, Darden, and Craver (2000) and recent process tracing methodology, the way one establishes that the links in purported social science mechanisms are causal is different from how one establishes that the links in biological mechanisms are causal[2]. The latter, MDC argue, relies on 'regularities', i.e. the recurrence of these links in other cases, whereas the former can rely on other evidence, e.g. counterfactual evidence as established in chapter 3. This argument thus provides a direct addition to recent discussions of whether regularities are necessary for mechanistic explanation (cf. Bogen 2005, 2008, Machamer 2004, Woodward 2002), by providing a case where they are not.

---

[1] A full overview situating MDC in the contemporary philosophical literature, and in the history of (natural) philosophy more generally, can be found in Nicholson (2012); a taxonomy of the new mechanist literature in particular can be found in (Andersen 2014a, b, Levy 2013).

[2] That is, in terms of Gerring's requirements for good research design I presented in the introduction, the way one establishes transparency is different.

My conclusion has two direct results. Firstly, MDC express the suspicion that their analysis "is applicable to many other sciences, and maybe even to cognitive or social mechanisms" (Machamer, Darden, and Craver 2000, 2). My conclusions about social mechanisms limit the scope of MDC's theory; they show to what extent MDC's new mechanist philosophy reaches beyond the mechanisms in their areas of expertise (molecular biology and neurobiology). Secondly, by analysing in what ways mechanistic reasoning in the social sciences differs from MDC's framework, I highlight what is meant by mechanistic reasoning in the social sciences, as opposed to what this term means in other areas.

This argument brings together the ideas contained in previous chapters, particularly regarding what constitutes good evidence for social processes (chapter 3) and how one conceptualizes the variables and causal relations in a social causal chain (chapters 4 and 5). In brief, my justification for the claim that regularities are hardly used as evidence of the genuine causal nature of the links of a social science mechanism, is that the causal heterogeneity of social science concepts throws a spanner in the works.

The chapter is set up as follows. I will begin the first part by discussing Machamer, Darden, and Craver's new mechanist theory in detail. In particular, I will consider their argument that in order to explain a phenomenon, the mechanism for that phenomenon must be based on regularities, and build on previous discussions of whether regularities are necessary for mechanisms by James Bogen, Peter Machamer, and James Woodward. After describing each part of MDC's theory, I discuss parallels and discrepancies between it and mechanistic reasoning in the social sciences. In particular, I show why the regularity view of mechanisms is not a good fit for the way mechanisms are explanatory in the social sciences, but argue that Machamer, Darden, and Craver's view on the discovery of mechanisms in the social sciences otherwise has direct parallels with recommendations in, amongst others, Andrew Bennett and Jeffrey Checkel's methodological guide *Process Tracing: From Metaphor to Analytic Tool* (Bennett and Checkel 2015).

In the second part of the chapter, in order to make my analysis more concrete, I illustrate my argument against the regularity view of mechanisms with a case study. I investigate Kenneth Schultz's study of the mechanism behind British use of coercive diplomacy during the Fashoda Crisis, an incident of territorial dispute that took place in Eastern Africa in 1898 between Britain and France and which led to a war scare but was eventually diplomatically resolved (Schultz 2001). In particular, I show that though the process Schultz discusses does not rely on regularities, as MDC require, it does rely on what is called 'productive continuity' by MDC.

There are two caveats with this chapter that I wish to address before continuing to the main line of argument. First of all, in this chapter I discuss under what circumstances a causal mechanism for a phenomenon (e.g. protein synthesis, or coercive diplomacy) can explain that phenomenon. I thus limit myself to considering what Daniel Nicholson has dubbed an 'epistemic conception' of causal mechanisms, that is, the conception of a causal mechanism as "a contingent explanatory description which heuristically abstracts away the complexity of a (…) system sufficiently to describe some localized causal process within it which leads to the realization of some function of interest" (Nicholson 2012, 160). Across philosophy of biology, the concept 'causal mechanism' is used in various other contexts (cf. Andersen 2014a, b, Levy 2013, Nicholson 2012), amongst which is the metaphysical thesis that "causal relations (…) exist in virtue of underlying mechanisms" (Levy 2013, 100)[3]. I will not go into these aspects of the mechanist position here.

_____

[3] Stuart Glennan advocates this position; he argues for a metaphysical picture of the world in which two events are causally related if and only if they are connected by a mechanism (Glennan 1996, 2002). I will not discuss this position further, except to note that in previous chapters I have already argued for a position that is incompatible with some readings of Glennan; there, I advocated the use of Woodward's interventionist view of causation as fundamental for process-based reasoning in the social sciences.

My second caveat is that though I believe my argument applies to the social sciences in general, I again take both my main example and the methodological framework I analyse from political science, as indeed I have done in the rest of the thesis. I believe that the conclusions I reach regarding both apply to other mechanistic studies and methods in the social sciences, given that (as I will discuss in more detail below) the heterogeneous nature of central concepts is not a problem unique to political science. Yet, my use of Schultz's study and Bennett and Checkel's methodological framework does rely on their usage of the term 'causal mechanism'. Though I have good reasons to believe their usage is representative for political science (reasons which I will detail below), if this usage is not illustrative for the social sciences as a whole, my conclusions may be limited to mechanistic reasoning in political science alone.

2. Machamer, Darden, and Craver's new mechanist philosophy

This section will describe in more detail MDC's claim that mechanisms for a phenomenon (e.g. protein synthesis) explain that phenomenon in virtue of the fact that their entities (e.g. DNA, amino acids) and activities (e.g. hydrogen bonding) are organized such that they are productive of regular changes. I will detail how MDC characterize the 'entities', 'activities', and 'regularities' they believe a mechanism consists of. I continue by discussing techniques for mechanism discovery as laid out in Darden and Craver (2002). At the end of every section, I draw the parallels and discrepancies between MDC's framework and how mechanistic reasoning applies in the social sciences.

2.1 MDC's characterization of causal mechanisms in biology

MDC characterize mechanisms as "entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions" (Machamer, Darden, and Craver 2000, 3). This characterization,

according to Lindley Darden at least, is not meant to be a definition, but rather is meant to capture how the term is used in biology, as informed by MDC's areas of research (neurobiology and molecular biology). Darden reveals that she has chosen to describe what biologists mean by the term 'mechanism', because she wants to provide some advice for how biologists may discover them (Darden 2008).

Machamer, Darden, and Craver give several examples from biology. One is the mechanism of chemical neurotransmission. In this mechanism, "a presynaptic neuron transmits a signal to a post-synaptic neuron by releasing neurotransmitter molecules that diffuse across the synaptic cleft, bind to receptors, and so depolarize the post-synaptic cell" (Machamer, Darden, and Craver 2000, 3). Chemical neurotransmission is regular in the sense that it happens between many different neurons; that, in turn, is the case because the separate activities that together make up the mechanism of chemical neurotransmission all "work always or for the most part in the same way under the same conditions" (Machamer, Darden, and Craver 2000, 3).

I will now analyse all the different parts of this characterization in turn: firstly, entities and activities, and in particular the associated notion of regularity; and second, the way mechanisms are described in biological practice, amongst which how one decides what to take as the start and finish conditions of a mechanism, and the idea of formulating a mechanism schema that is then filled in for each instantiation of the mechanism.

2.2 Entities, activities, and productive continuity

2.2.1 In biology

In molecular biology and neurobiology, examples of entities include such things as cell walls and DNA. Entities, simply speaking, are *things*. They have a spatiotemporal location, and we can distinguish them in a straightforward way from other entities (e.g. because they are divided from other entities by a boundary,

such as a membrane, are internally bound together and disconnected from other entities, or because they consist of different chemicals). Sometimes, entities can be tracked through a developmental or evolutionary history. Note that MDC do not wish to define any of a mechanism's parts (i.e. the entities, activities, etc.) "by giving necessary and sufficient conditions. Rather, the goal is to provide strategies that may serve as useful guides for finding [them]." (Darden 2008, 961)

Entities are interdependent with the other part of a mechanism in MDC's framework, the activities. Examples of activities include hydrogen bonding, transport, and depolarization. The properties of the entities constrain what kind of activities they can undertake, and the kinds of activities undertaken constrain what kinds of properties the entities can have. For instance, "[t]he neurotransmitter and receptor, two entities, bind, an activity, by virtue of their structural properties and charge distributions" (Machamer, Darden, and Craver 2000, 3). In MDC's framework, activities are seen as types of causes: they "are producers of change; they are constitutive of the transformations that yield new states of affairs" (Darden 2008, 962). So, for instance, in the mechanism behind penicillin's effect on pneumonia, "[i]t is not the penicillin [the entity] that causes the pneumonia to disappear, but what the penicillin does [the activity]" (Machamer, Darden, and Craver 2000, 6).

Entities and activities together are linked in what MDC call 'productive continuity':

> Complete descriptions of mechanisms exhibit productive continuity without gaps from the set up to termination conditions. Productive continuities are what make the connections between stages intelligible. If a mechanism is represented schematically by $A \rightarrow B \rightarrow C$, then the continuity lies in the arrows and their explication is in terms of the activities that the arrows represent. A missing arrow, namely, the inability to specify an activity, leaves an explanatory gap in the productive continuity of the mechanism. (Machamer, Darden, and Craver 2000, 3)

2.2.2 In the social sciences

What are the entities and activities of social mechanisms, and how can we best describe their particular kind of productive continuity? As we have seen in chapter 3, the term 'causal mechanism' is used to explain, for instance, the influence of transnational insurgents on local insurgents' radicalization of tactics in Bakke's study of the Second Chechen War. It is important to note here, again, that there are some important differences between the way the term 'causal mechanism' is used in biology, and the way it is used in social science.

Firstly, as Daniel Little stresses, it is important not to take the analogy between physical mechanisms and social mechanisms too far: in the social sciences, mechanisms are an 'umbrella term', "a 'family resemblance' term that captures a number of different instances of collective behaviour and agency" (Little 2011, 277).

Secondly, as I argued in chapter 3, process tracers do not study mechanisms directly: instead, they trace the causal chains connecting a putative cause and effect of interest (which result from combinations of mechanisms and background conditions). In biology, we may (according to MDC) see a straightforward relation between one mechanism and the process observed in a particular instance, e.g. the mechanism of protein synthesis and the production of such a protein in a particular cell under study. In the social sciences, things are less straightforward; think, for instance, of the chain of events in Chechnya, which Bakke claimed was the result of different interacting mechanisms (including relational diffusion and brokerage) and favourable background conditions.

Thirdly, things are further complicated because the entities and activities of social mechanisms must be constrained more generally by the principles I set out for social phenomena in chapter 2. There, I showed that social science phenomena have a wide variety of associated meanings, and accordingly they can be systematized in a variety of ways. Moreover, many social science terms pick out a group of phenomena that cannot be circumscribed by a set of necessary and sufficient (boundary) conditions. Which systematizations and boundary conditions we allow depends on the causal structure of the phenomena, I argued there. Taking these

insights on board means that the units in social science mechanisms will not be as clearly defined as biological entities: we cannot easily rely on a spatiotemporal location or set of boundaries to distinguish social 'entities', and similarly social activities are not as straightforwardly defined as they are in biology.

What, then, about productive continuity? Here, I would argue, it is important to refer back to the discussion of process tracing in chapter 3. There, I showed how one can give evidence for the causal nature of (some) links of the causal chain connecting a putative cause and effect using Woodward's interventionist theory of causation. Productive continuity in the social sciences, I believe, is most fruitfully considered in terms of the process that process tracers are tracing: that is, in terms of the causal nature of the links in a causal chain of interest.

2.3 The regularity debate

2.3.1 As applied to biology

In MDC, mechanisms only explain because the activities in these mechanisms are regular; by this, MDC mean that the activities "work always or for the most part in the same way under the same conditions" (Machamer, Darden, and Craver 2000, 3). In other words, the same activity working in some other biological system but in the same context will produce the same effect.

In the areas of biology that MDC have studied, regularities like '*A* causes *B*' are one of the prime sources of evidence for counterfactuals like 'had *A* not happened, then *B* would not have happened'. Thus, MDC claim that counterfactuals like "if this single base in DNA were changed and the protein synthesis mechanism operated as usual, then the protein produced would have an active side that binds more tightly" (Machamer, Darden, and Craver 2000, 8) are supported by what we know about the regularities in the protein synthesis mechanism. If we want to find out what the effect of a change in that particular single base in DNA would be on the tightness of the bond in the protein, knowledge about the regular causal relation between that base and that bond (e.g. in other cells) will be helpful.

Whereas authors like William Bechtel (cf. Bechtel and Abrahamsen 2005, Bechtel and Richardson 1993) refer to laws to found these regularities on, MDC claim any reference to laws is superfluous. They believe that "no philosophical work is done by positing some further thing, a law, that underwrites the productivity of activities" (Machamer, Darden, and Craver 2000, 8). James Woodward agrees with MDC that in many cases in biology, "the relationship governing the intermediate steps of this process are too local and too susceptible to exceptions to count as plausible candidates for laws" (Woodward 2002, S373). In order to detail what else might be behind the 'regularities' in MDC's characterisation, Woodward has suggested understanding 'regular productive behaviour' in terms of his interventionist theory of causation, which I discussed in detail in chapter 3.

However, contrary to my framework in chapter 3, both Woodward and the original MDC text characterize activities in terms of generalizations, i.e. as 'types of causes that describe something that acts in the same way under the same conditions'. Woodward then cashes out these generalizations using his interventionist theory. However, it is this *generalizability* that leads to difficulty when it comes to considering an area of research like political science, where there are no 'causally homogeneous', 'stable' activities across a range of cases. We may think that neurons act in a similar way under the same conditions, but there is no reason to believe that (to give an example direct from Schultz's case study, which will be the topic of section 3) a democratic government will act in a similar way unless the same conditions are spelled out in excessive detail. I earlier mentioned that MDC's theory is often described as part of a 'new mechanist' tradition in philosophy of biology. This tradition is broadly characterized by an emphasis on explanation in terms of mechanisms as opposed to, amongst others, the deductive-nomological model of explanation (cf. Andersen 2014a). To discuss this problem with the scope of MDC's theory more thoroughly, I wish to outline another new mechanist's alternative to MDC's emphasis on regularities.

MDC's view that regularity is necessary for mechanistic causation is not accepted by every new mechanist. James Bogen suggests that although entities and activities are indeed organized so that they are productive of changes, these changes do not

need to be regular (Bogen 2005, 2008). After all, Bogen argues, there exist irregularly operating and stochastic mechanisms, as well as mechanisms that operate just once. In later work, Machamer has taken Bogen's side (Machamer 2004, 37, f1).

Bogen's view then is a slimmed down version of MDC. He does not rely on regularities, but simply on what MDC call 'productive continuity'. In MDC's framework, "[a mechanism's] regularity is exhibited in the typical way that the mechanism runs from beginning to end; what makes it regular is the productive continuity between stages" (Machamer, Darden, and Craver 2000, 3). However, in Bogen, this is not the case. In short, his argument is as follows. Though MDC believe that "generalizations describing natural regularities are essential components of causal explanations" (Bogen 2005, 397), Bogen argues that "causal productivity and regularity are by no means the same thing" (Bogen 2005, 397).

Another way of interpreting this argument is as follows. Although the existence of regularities is one source of evidence for the causal productivity of a mechanism (since such regularities provide evidence that the link between some parts of the mechanism is genuinely causal), this is not the *only* source of evidence for productive continuity and indeed it would be misleading to think so, since excluding other sources of evidence means we cannot consider singly recurring mechanisms as genuinely causal.

Note that Darden argues against Bogen, not on the philosophical grounds discussed above, but because she believes Bogen's suggestion does not accurately reflect molecular and neurobiological practice. I take this to be a descriptive claim, not a normative claim; moreover, it does not distract from the argument above that there are *other* sources of evidence for productive continuity *besides* regularities. Moreover, this descriptive fact about molecular and neurobiological practice does not lend credit to the much stronger claim that mechanisms are explanatory *only* because their activities are regular, as MDC sometimes seem to imply. After all, just because in molecular and neurobiological practice scientists do not or hardly ever refer to other sources of evidence is not to say that such evidence cannot fruitfully be used.

There is a second, related aspect to the regularity of biological mechanisms. Not only do MDC claim that regularities give us evidence that the parts of the chain of the mechanism are genuinely causal, i.e. that the fact a mechanism's links recur in more than one case gives us evidence of the causal nature of those links, they also claim that the mechanism *as a whole* recurs. According to Darden, 'regularity' also covers "the typical way [in which] the mechanism runs from start to finish" (Darden 2008, 964). I will come back to this below.

2.3.2 As applied in the social sciences

Let me now apply this evidential argument against the regularity view of mechanisms to social science practice. I will then come back to Bogen's interpretation and evaluate whether it can describe what is going on in the social sciences.

In this section, I will argue that in the social sciences, regularities are not the only source of evidence for the productive continuity of social mechanisms (i.e. for the claim that the links between the elements of the chain connecting a putative cause and effect of interest are genuinely causal).

This will be relatively straightforward if we take into account my conclusions in chapter 3. There, I argued that we should give evidence of interventions to support the links of a casual chain, i.e. what MDC call 'productive continuity'. In chapter 3, I argued that to give evidence for the causal nature of the links $Z_i \rightarrow Z_j$ in a chain between a putative cause X and effect of interest Y, we must find intervention variables I for the relation $Z_i \rightarrow Z_j$. These intervention variables then support the counterfactual claim that if Z_i had not occurred (or had some other value), Z_j would not have occurred (or would have had a different value). Intervention variables can be based on actual interventions, natural experiments, or thought experiments.

Sometimes, interventions are based on counterfactual reasoning for one case alone, and not based on regularities. Recall, for instance, Bakke's argument in chapter 3 that transnational insurgents in Chechnya induced the radicalization of tactics

between the First and Second Chechen War. We saw there that we had to, for instance, provide evidence for the counterfactual 'the Chechen insurgents would have used suicide terrorism less in the Second Chechen War if they had been prevented from being shown suicide bombing videos by the transnational insurgents'. This, I admitted, required some comparisons (between the Chechens before the transnational insurgents arrived, and the Chechens after the transnational insurgents arrived) to see whether their previous reluctance to use suicide tactics could be extrapolated to show that without outsider influence, the Chechens would not have resorted to such tactics. However, this is not based on a regularity, i.e. it is not based on evidence that the same connection between watching videos and developing tactics applied in other conflict situations.

Moreover, as I have discussed earlier in this thesis, there are other reasons to be mistrusting of the replicability of (parts of) chains of events in the social sciences. In particular, I have argued that in order to generalize from one case study to a general causal mechanism, one either has to find evidence for a causal homogeneity condition, or needs to assume epistemic homogeneity. Recall that the causal homogeneity condition asks of us that the influence of the cause on the effect is, *ceteris paribus*, equal for all cases. Recall also that epistemic causal homogeneity means conceptualizing a putative cause and effect of interest in such a way that we assume homogeneity. Without ontological homogeneity, it is hard to see how one can get the regularity that MDC require of the activities in mechanisms. So, mechanisms in the social sciences do not meet MDC's demand that the activities are *regular*. Another way of putting this would be to agree with Bogen's claim above: we are often faced with causal links of a chain which occur (in Bogen's words) irregularly, stochastically, or only once.

This argument, then, restricts the scope of MDC's theory: in biology, the evidence for the intervention variables supporting a counterfactual like "if this single base in DNA were changed and the protein synthesis mechanism operated as usual, then the protein produced would have an active side that binds more tightly" (Machamer, Darden, and Craver 2000, 8), may often be in terms of regularities. Knowing how the DNA base and the protein bond relate in a variety of other

154

contexts will lend support to this counterfactual. Yet in the social sciences, getting evidence for productive continuity from regularities is unlikely. Admittedly, the counterfactual evidence one collects in social science can rely on cross-case comparisons, as I outlined in chapter 3 in the discussion of Woodward's interventionism. Yet it seems a stretch, given what I have argued in chapters 4 and 5, to call such cross-case comparisons 'regular' in the same way that the relation between a DNA base and protein bond is regular.

2.4 Describing mechanisms

2.4.1 In biology

As argued, the new mechanists believe that one can explain a phenomenon by providing the mechanism for that phenomenon (cf. Wimsatt 1972, 67). Discovering a mechanism for a phenomenon consists of several stages.

Describing a mechanism is to some extent perspectival (Darden 2008). One good way of seeing this is by looking at the start or set-up conditions and finish or termination conditions. These conditions, MDC claim, are idealized descriptions of the beginning and endpoint of the mechanism, respectively. What we take to be the beginning and endpoint of the mechanism is context-dependent. Set-up conditions (which include the relevant entities and their structural properties, spatial relations, and orientations, as well as any enabling conditions for the subsequent working of the mechanism) may be the result of prior other mechanisms that we choose to ignore. Similarly, what we choose to regard as the endpoint of the mechanism (e.g. a state of rest or equilibrium, or of repression or activation, elimination or production of something) depends on what we are interested in.

A researcher most likely starts with an incomplete model for the mechanism. In such a 'mechanism sketch', fundamental entities and activities might well be missing. The mechanism may in other words contain gaps in its 'productive continuity': whole stages of the mechanism may yet be unknown. Those entities or

activities must be found in order to definitively show a causal link from the mechanism's start to its finishing conditions. Once all the gaps in the productive continuity of the sketch are filled, the scientist has produced what MDC call a 'mechanism schema', i.e. "a truncated abstract description of a mechanism that can be filled with descriptions of known component parts and activities" (Machamer, Darden, and Craver 2000, 15).

2.4.2 In the social sciences

Bennett and Checkel make clear that the start or set-up and finish or termination conditions in process tracing "depend on how a researcher defines the puzzle or question they are trying to explain" (Bennett and Checkel 2015, 26) and that "[e]ven within one well-defined research question, the proper starting point can be subject to debate" (Bennett and Checkel 2015, 26). However, they also believe that some start and finish conditions will be more useful than others. For instance, a reasonable place to start could be "a critical juncture at which an institution or practice was contingent or open to alternative paths, and actors or exogenous events determined which path it would take" (Bennett and Checkel 2015, 26). Bennett and Checkel argue that researchers should not only consider which moments in a chain of events were critical, but also which moments had the potential to be critical but were not, i.e. they should also consider "times at which institutions could have changed, perhaps due to some exogenous shock, but did not." (Bennett and Checkel 2015, 27)

2.5 Filling the gaps: forward chaining, backtracking, and schema instantiation

2.5.1 In biology

How does a researcher go from an incomplete sketch to a filled-in schema? According to later work by Darden and Craver, "there are more or less reliable, yet

inherently fallible, strategies for discovery" (Darden and Craver 2002, 19). Darden and Craver describe two techniques for doing so, 'forward chaining' and 'backtracking'. Forward chaining and backtracking make use of what we already know about particular parts of the mechanism (its entities and activities and their properties) in order to fill in preceding or consequent steps in that mechanism. The techniques are possible because of the interdependency between entities and activities: "[e]ntities and a specific subset of their properties enable the activities in which they engage (given appropriate conditions). Furthermore, activities require distinct types of entities and properties of those entities as the basis for such acts." (Darden and Craver 2002, 21-22) Therefore, "[l]ooking forward, each stage [of the mechanism] must give rise to, allow, drive or make the next. Conversely, looking back, each stage must have been produced, driven or allowed by the previous stage(s)." (Darden and Craver 2002, 4)

There are several subtypes of forward chaining and backtracking: considering (1) activity-enabling properties of entities (e.g. spatial and structural properties, charges, valences); (2) activity consequences ("what is expected of the entities in the subsequent stage, given the prior occurrence of some activity", (Darden and Craver 2002, 23)); (3) activity signatures of entities ("how such entities could have been produced or what activities could have given rise to, driven, made or allowed this later stage" (Darden and Craver 2002, 24)); and (4) entity signatures of activities ("the characteristic features of an activity [that] may provide clues as to the entities that engaged in it" (Darden and Craver 2002, 24)).

Once the mechanism schema is produced, the researcher may still have to articulate how the parts and activities are instantiated in a particular case under consideration, thus having to make the schema less abstract. Darden and Craver call this 'schema instantiation'. During schema instantiation, we find 'black boxes' in the schema that ought to be filled in with the entities and activities of a concrete scenario in which the mechanism is at work (Darden and Craver 2002, 20).

MDC stress that mechanisms in molecular and neurobiology are often 'multilevel'. The example they give is of the central nervous system: "the activation of the

sodium channel is a component of the mechanism of depolarization, which is a component of the mechanism of chemical neurotransmission, which is a component of most higher-level mechanisms in the central nervous system" (Machamer, Darden, and Craver 2000, 13). One mechanism can thus be a component of another, in what MDC call a "part-whole hierarchy" (Machamer, Darden, and Craver 2000, 13). What we consider to be the 'lowest' level mechanisms in this hierarchy (e.g. whether we are happy to consider the mechanism of chemical neurotransmission as the lowest, or wish to know in more detail how its constitutive mechanism of depolarization works) is context-dependent. In MDC's characterization of mechanisms, the 'lowest level' mechanism is simply the lowest level that the scientific field under study is still constructing mechanisms for. If neurobiologists are typically uninterested in the workings of the sodium channel itself, philosophers of neurobiology may call the mechanism of depolarization the 'lowest level'.

The fact that mechanisms like depolarization can themselves be simply steps of a 'higher-level' mechanism like chemical neurotransmission has an interesting connection with the distinction I made at the end of section 2.3.1, where I argued that not only the parts of the biological mechanism are regular, the mechanism as a whole can be regular too. This seems necessary in a case where this mechanism is part of a hierarchy of mechanisms.

2.5.2 In the social sciences

We have seen in section 2.3.2 that social science mechanisms display productive continuity, but that evidence for this continuity is not necessarily based on regularities. Can social science nevertheless benefit from the discovery techniques of schema instantiation, forward chaining, and backtracking? Do process tracers ever attempt to fill gaps in the productive continuity of the mechanism they are studying? More generally, is regularity *required* for successful schema instantiation, forward chaining, and backtracking?

Based on my arguments in earlier chapters, I would argue regularity is not required. Forward chaining and backtracking are, under those terms, discussed in

methodology texts like Bennett and Checkel (2015). Andrew Bennett himself argues elsewhere that "[t]here is no guarantee that researchers will include in their analyses the variable(s) that actually caused *Y*, but process tracing backward from observed outcomes to potential causes – as well as forward from hypothesized causes to subsequent outcomes – allows researchers to uncover variables they have not previously considered" (Bennett 2010, 209).

Moreover, just like Machamer, Darden, and Craver argue that part of the mechanistic project means filling in the 'black boxes' of a mechanism schema for the particular case under study, Bennett and Checkel also argue that one ought to operationalize a mechanism stated in more general terms for a particular case under study. "We cannot stress enough that theories are usually stated in very general terms; they must therefore be operationalized and adapted to the specific processes in particular cases (…)" (Bennett and Checkel 2015, 30). We have already discussed this in chapters 3 and 4 when we talked about the observable implications of more general mechanisms, which are compared to causal process observations (CPOs).

In the next section, I will show how one researcher in particular discovered a social mechanism. This will then illustrate how, despite an absence of regularities to fall back on, social scientists can nevertheless move forward.

3. Example: Schultz's discussion of the Fashoda Crisis

In this section, I will illustrate my comparative analysis above by discussing a representative example of mechanistic reasoning from political science, i.e. Kenneth Schultz's analysis of the events of the Fashoda crisis (Schultz 2001). I will detail what, if anything, can be considered the entities and activities for this chain of events. I will then show why it is unreasonable to assume that they are 'regular'. I finish by concluding that despite not being regular, Schultz's mechanism is explanatory.

3.1 How representative is Schultz for mechanistic reasoning in the social sciences?

Before I discuss the case study, it is worth asking to what extent Schultz's approach is representative for mechanistic reasoning in the social sciences more generally. It is worth keeping the following caveats in mind whilst reading the rest of this chapter. Firstly, Schultz is a political scientist, and so if anything he is representative of mechanistic reasoning in that discipline in particular. (Whether or not his analysis is representative of, say, analytical sociology is an entirely different matter, and I will leave this, to speak with MDC, 'an open question'.) Secondly, Schultz's work has been hailed by several prominent methodologists as 'convincing' (Bennett 2010, Bennett and Checkel 2015, George and Bennett 2005). Several authors have analysed Schultz's account of the Fashoda crisis. In the 2005 book *Case Studies and Theory Development in the Social Sciences*, George and Bennett express that Schultz's work is "generally rigorous", "well done", and his case studies are "convincing" (George and Bennett 2005, 56). Lastly, Schultz's work is used as a teaching example (Collier 2011a) aimed at early-career political scientists who wish to learn process tracing.

3.2 Schultz's argument

Kenneth Schultz is interested generally in how democracies use coercive diplomacy, i.e. how democracies threaten an opponent with the use of force to make them stop or change certain behaviour. Schultz claims that the governments of democratic nations can more effectively use coercive diplomacy to triumph in international crises than non-democratic governments can. Schultz postulates a general causal mechanism to explain why being in a democracy has this effect: this mechanism is called 'signalling'. In brief, his argument is that "states which permit public competition for political office can generate more credible signals of resolve than can states in which competition is suppressed or private" (Schultz 2001, 162), and thus, "public signals of domestic support can increase the credibility, and hence the

effectiveness, of threats made by democratic governments." (Schultz 2001, 197) Domestic support shows the opponent that the government is in all likelihood not bluffing; they are prepared to actually go through with the threat.

Moreover, governments are more likely to succeed if their threat to use force is supported by the opposition parties than if it is opposed. If a threat to use force is supported by the opposition parties, this leads to what Schultz calls the 'confirmatory effect': "a signal sent by two actors with opposing interests is more informative than a signal sent by one actor with known incentives to misrepresent its preferences." (Schultz 2001, 162) Simply speaking, although a government on its own may have good reasons to bluff about their threat of force, an opposition party does not have good reasons to bluff when it comes to supporting the threat. Rather, Schultz argues that the opposition party will only support the threat if this threat genuinely has a chance of succeeding; only then is such support in the party's electoral interests. This means that a threat to use force that is supported by the opposition parties is taken more seriously by the opponent.

The main part of Schultz's book is devoted to developing a game theoretic model of democratic countries' use of information and signalling in coercive diplomacy during international crises. Schultz tests this model with statistical evidence and finds a positive correlation between support by the opposition parties on the one hand and 'deterrence success' on the other, even after controlling for military capabilities and indicators of issue salience. Moreover, he also finds that if the opposition party does not support the threat of force, the probability of that threat's success is similar to the probability of a nondemocratic government's threats being successful.

Schultz devotes a (smaller) part of his book to a process-tracing study of opposition party support for a democratic country's war threat, the Fashoda crisis[4]. This is the

---

[4] I will only focus on Schultz's case study into whether the causal mechanism postulated in his model was present during the Fashoda Crisis, and not on his statistical analysis. Note that Schultz does not use his statistical analysis to give

part that, as I outlined above, is critically acclaimed and used by many methodologists as an example of good process tracing (cf. Bennett 2010, Bennett and Checkel 2015, Collier 2011b). Schultz believes that the Fashoda crisis is representative of his more general model because the United Kingdom, one of the nations in the crisis, had a political system that is very similar to the way his game theoretic model represents domestic politics (cf. Fearon and Laitin 2011, Wimmer, Cederman, and Min 2009). Let me now outline the Fashoda case very briefly, before turning to Schultz's process tracing in that case.

3.3 The Fashoda Crisis case study

In the last decades of the nineteenth century, most of Africa was divided by the major European colonial powers. The Upper Nile region, in what is now Sudan, was the exception, having been evacuated by the Egyptians in 1884. The British powers claimed that the Upper Nile region still belonged to the Egyptians, but the French claimed it had no official owner. In 1895, the British Foreign Office warned that any attempt by the French to claim the region would be considered an unfriendly act. The French, however, sent Major Jean Baptiste Marchand to the region to claim it, and after a two-year overland journey from the Congo, Marchand arrived with a small French force in Fashoda, in the Upper Nile Valley, on July 10 1898. The British, in response, had sent General Herbert Kitchener, who arrived in Fashoda on September 18 1898, with a larger Anglo-Egyptian army. The meeting between Marchand and Kitchener was courteous, whilst the British and French governments back home in Europe attempted to settle the matter. The British

evidence for any of the causal links of the chain of events during the Crisis; therefore, I would argue that any criticism that Schultz's statistical work lends evidence for MDC's emphasis on the regularity of mechanisms' links is unfounded. It does, however, give some evidence that social mechanisms can as a whole be recurring. I will come back to this issue in the conclusion of the thesis.

threatened to go to war with France over the Fashoda region, and eventually, after two months, the French backed down. Marchand left Fashoda to the Anglo-Egyptian army. (cf. Chamberlain 2013)

In his analysis of the Fashoda crisis, Schultz first outlines and rejects three alternative hypotheses about the events at Fashoda (the balance of power theory, democratic peace theory, and institutional constraints argument). The balance of power theory explains the crisis by pointing out that the British forces were much stronger than the French. However, Schultz argues, this cannot explain all aspects of the crisis: the French force was so small that the French "had to know from the outset that, if a confrontation arose, Marchand would not be able to hold his own militarily" (Schultz 2001, 178). So why, if the balance of power theory is true, did the crisis occur in the first place, and why did it last so long? (Schultz 2001, 177-180)

Both the democratic peace theory and institutional constraints arguments claim that the crisis de-escalated because, respectively, democratic norms or institutions made the British and the French restrain themselves. However, as Schultz shows, the British were hardly restrained; rather, their rhetoric and actions were aggressive (Schultz 2001, 180-183). For instance, the British were well aware of the Dreyfus affair's destabilizing influence on the French government. This affair was occurring at the same time as the Fashoda crisis was unfolding, and although the British were aware that their position on Fashoda was worsening the French political situation, they did not moderate their position in response.

After rejecting these alternative hypotheses, Schultz then offers his own hypothesis, which applies the signalling mechanism he has set up in earlier chapters to the Fashoda incident. He argues that the French decision to lay claim to the region was based on the wrong impression of British intentions, in particular the belief that the British would prefer making concessions to fighting a war. However, the British government's mind seemed made up. This signal was especially clear after they made public a 'blue book' which contained all the exchanges between Britain and France up until that point. This, Schultz shows, meant that the British public read the uncompromising position of the government first-hand. Thereby, "[British

prime minister] Salisbury effectively painted himself into a corner: retreat from this position would entail substantial political costs" (Schultz 2001, 187).

Only when the British ultimately convinced the French their threat was genuine, the French backed down:

> Over the course of the following month, the French learned that [their] initial expectations had been incorrect. The signals emanating from Britain during this period took many forms, but all contained the same message: the French would have to evacuate Fashoda unconditionally or face war. Actions taken by the government – and confirmed by the opposition – made it clear where the former's political incentives lay. Public opinion was such that the political risks of war were small while the political risks of compromise were potentially large. (Schultz 2001, 186)

Thus, Schultz claims that signalling is the main mechanism behind the success of British coercive diplomacy during the Fashoda incident:

> Although the balance of military power was important, the decisive factor was Britain's ability to convince France that it was willing to use that power rather than make the slightest concession. In this, the government was greatly aided both by its ability to generate large audience costs[5] and by the confirmatory signals that emerged from opposition figures of all stripes. French leaders, by contrast, generally avoided actions which would tie their hands and saw their position undercut by vocal opposition to the use of force among large segments of the body politic. By the end of the crisis, France knew that Britain was ready to fight, and Britain knew that France was not. (Schultz 2001, 195)

---

[5] Audience costs are domestic political costs that leaders suffer if they escalate an international crisis, and then back down (cf. Fearon 1994).

3.4 Analysing the Schultz case study through MDC's framework for causal mechanisms

As we have seen, Schultz's mechanism is 'signalling'. He tests whether this mechanism explains the French decision to back down at the end of the Fashoda incident. Hence, we could consider his model of signalling (outlined in the first chapters of *Democracy and Coercive Diplomacy*) as a mechanism schema, which is subsequently filled in for the particular case of Fashoda. Of considerable interest here is what, in this case study, would be the equivalent of MDC's entities, activities, and productive continuity.

Let us sum up the Fashoda crisis example. Here we have:

Activities	Forming expectations (by the French government); producing a blue book (by Salisbury and the British government); interpreting signals (by the British public, the French government, etc.)
Entities	Sir Herbert Kitchener; Jean Baptiste Marchand; the British government; the British public; the British press; etc.
Mechanisms	Signalling (defined in terms of Schultz's game theoretic model)
Start conditions	The colonial powers' division of Africa by the end of the nineteenth century; British electoral interests; French electoral interests
Finish conditions	The French decision to back off
Counterfactuals	e.g. 'If the British opposition parties had not sided with the British government in their war threat against France, France would not have backed off'

3.4.1 Regularities in Schultz?

Let us now turn to a discussion of whether the productive continuity in the chain of events Schultz discusses is supported with evidence from regularities. For this purpose, it is useful to refer to Andrew Bennett's analysis of the case study in terms more often used in process tracing methodology: 'smoking gun' and 'hoop' tests. Smoking gun tests, in brief, are tests that are sufficient (but not necessary) for confirming a causal claim; hoop tests are tests that are necessary (but not sufficient) for confirming a causal claim (cf. Collier 2011b). Failing a hoop test refutes a causal hypothesis, and succeeding in a smoking gun test confirms it. Thus, one would like one's own hypothesis to meet smoking gun tests, while in the meantime one would like to see alternative hypotheses fail hoop tests.

Bennett analyses Schultz's case study as follows:

> Schultz supports [his] explanation with *smoking gun* evidence. The credibility of Britain's public commitment to take control of the region was resoundingly affirmed by the opposition Liberal Party leader Lord Rosebery (…). Meanwhile, France's Foreign Minister, Theophile Delcasse, initially voiced an intransigent position, but his credibility was quickly undermined by public evidence that other key French political actors were apathetic toward, or even opposed to, a war over Fashoda (…). Within a matter of days after such costly signalling by both sides revealed Britain's greater willingness and capability to fight for the Upper Nile, France began to back down, leading to a resolution of the crisis in Britain's favour. In sum, the close timing of these events, following in the sequence predicted by Schultz's theory, provides *smoking gun* evidence for his explanation (…). (Bennett 2010, 212)

Smoking gun evidence of a researcher's own hypothesis alone is not enough to weaken the plausibility of *alternative* hypotheses in the literature, Bennett claims; however, Bennett interprets Schultz's analysis of the alternative hypotheses (the balance of power theory, democratic peace theory, and institutional constraints argument) as evidence that these hypotheses fail hoop tests. With the alternative

hypotheses thus rejected (after all, they have failed the necessary tests) and his own hypothesis thus supported (after all, it has met a sufficient test), Bennett deems "Schultz's explanation of the Fashoda case convincing" (Bennett 2010, 212).

It thus seems to be the case that the Fashoda case study in Schultz's work is not seen as convincing by peers like Bennett because Schultz gives evidence for the links between the chain of events in terms of regularities (e.g. 'whenever a blue book is produced on a decision in crisis, public opinion will prevent the government from backing down'), except in the most commonplace sense that Schultz assumes e.g. that public announcements are regularly heard and understood by the public. Instead, what makes the story powerful is the fact that Schultz discounts his opponents' alternative hypotheses using so-called 'hoop tests', and supports his own hypothesis with causal-process observations which are so called 'smoking gun tests'. Moreover, the chain of events Schultz traces shows great continuity. For these reasons, Bogen's view of what constitutes a causal mechanism (productive continuity, but not necessarily productive continuity supported by regularity) is more fitting for the social sciences than MDC's framework.

Admittedly, there is a generalization in Schultz's work in the following sense: the Fashoda case study is meant to illustrate the higher-level theory that Schultz has postulated in earlier chapters, i.e. the theory which shows that if there is domestic support for a threat made as part of coercive diplomacy in an international crisis, that threat is more likely to lead to the opponent backing off. First, Schultz tries to find statistical evidence for this theory; then, he provides process-tracing, singular case study evidence for the theory. This, however, does not mean that the regularity he has hypothesized is used to give evidence for the individual causal links in the Fashoda case study.

4. Conclusion

In this chapter, I have compared Machamer, Darden, and Craver's new mechanist framework, which analyses causal mechanisms in terms of activities, entities, and

regularities, to causal mechanistic reasoning (process tracing) in political science. I showed that whilst the causal mechanisms in MDC's framework have many parallels with social mechanisms, amongst others in terms of the way these mechanisms are discovered through their productive continuity, social mechanisms have no straightforward equivalent to the regularities required by MDC. I illustrated this by means of the case study of the Fashoda crisis in Schultz's study of coercive diplomacy, in which the productive continuity of the mechanism 'signalling' does not rely on regularities but rather is proven by means of smoking gun tests that rely on the close timings of the events of the crisis. By showing that not all evidence for productive continuity takes the form of regularities, my chapter has thus made an addition to recent discussions between Machamer, Woodward, and Bogen, of whether regularities are necessary for mechanistic explanation, by providing a case where they are not.

To conclude, what are we to make of MDC's suspicion that their analysis "is applicable to many other sciences, and maybe even to cognitive or social mechanisms" (Machamer, Darden, and Craver 2000, 2)? The above shows that MDC's theory as it currently stands may provide a good description of molecular and neurobiological practice, but is as yet insufficient to describe certain aspects of social scientific practice. What is needed is a broader understanding of productive continuity, which respects alternative sources of evidence for the links of a causal chain. Drawing comparisons between the conclusions of this chapter and the conclusions of earlier chapters, this broader framework will have to deal with cases of singular causation and cases in which one only has evidence for the epistemic homogeneity of a causal relation. Only then can MDC's theory reach beyond the mechanisms in molecular biology and neurobiology.

Chapter 7

Conclusion

In this thesis, I have discussed how we can pursue good causal inquiry in the social sciences, despite the fact that no two instances of social phenomena are entirely alike, causally speaking. Think back, for instance, to civil war: every conflict we call a civil war in everyday language has its own history and development, its own major actors, and its own political landscape. Yet, there are occasions when we wish to build upon the non-technical usage of a term like 'civil war' and develop the background concept into a concept systematized for academic study, in order to learn more about the causes of its instantiations. On those occasions, we compare or generalize over subtly different cases, and thus we have to simplify: we must ignore some causally relevant differences between the cases. I argued that causal inquiry can be fruitful despite these limitations: by carefully outlining and continually challenging our simplifying assumptions, we can learn more about the causes of such complex phenomena as civil war. I focused my attention on one particular social scientific methodology, process tracing, and on one particular social scientific discipline, conflict studies. This focus has allowed me to detail how causal inquiry, comparison, and generalization can work despite the limitations above.

In my introduction, I promised this thesis would shed light on two main aspects of social science research design: *transparency* (the demand that the research method gives an insight into the process connecting cause and effect) and *comparability* (the demand that the samples under study in a research design are similar enough to draw conclusions). In this conclusion I will investigate, on the basis of the research presented in this thesis, whether current best practice in process tracing meets these demands on good research design, and to the extent that it does not, how we might improve the methodology on the basis of my research.

1. Transparency

Transparency, the demand that the research design offers evidence about the process connecting cause and effect, was the topic for chapters 3 and 6. I presented two ways of analysing process tracing, one based on James Woodward's interventionist theory of causation, and one based on the new mechanist theory first proposed for molecular and neurobiology by Machamer, Darden, and Craver.

In chapter 3, I set out a formalization of process tracing. I showed that (top-down) process tracers formulate a hypothesis about what may be the cause of some effect they are interested in, and by what mechanisms the two could be connected. Subsequently, they try to provide support for this hypothesis in a case study, as well as refute any existing rival hypotheses in the literature. In the simplest case (in which the researcher only postulates one mechanism), I formalized process tracing as follows: a researcher holds that some causal mechanism Z is behind a process linking a putative cause, X, and the observed effect, Y. This mechanism has observable implications, i.e. the traces of a set of intervening variables Z_i such that $X \rightarrow Z_1 \rightarrow Z_2 \rightarrow \dots \rightarrow Y$ (where $Z_i \rightarrow Z_j$ means that Z_i causes Z_j). It is this chain of events that process tracers trace.

Current best practice in process tracing, I showed, states that if we wish to establish that X is a cause of Y, it is sufficient to observe the deductive implications of the purported intervening variables of the mechanism in a case study. However, I argued that we must do more than this: we ought to also show that all links $Z_i \rightarrow Z_j$ of the chain connecting X and Y are genuinely causal. Using Woodward's manipulability theory, I argued that to show that the relations hypothesized are genuinely causal we would have to provide evidence that there is a possible intervention for each link of the chain. I argued that one way of gaining knowledge of intervention variables is by comparing and contrasting one case to another, either with a natural or a hypothetical experiment, which in both cases requires a sophisticated analysis of to what extent such cases are similar. Thus, we may conclude that current best practice in process tracing does *not* meet the standard of

transparency; without further evidence that the steps of the chain are genuinely causal, process tracing is not transparent.

In Chapter 6, I approached the question of transparency from a different perspective. I discussed Machamer, Darden, and Craver (MDC)'s characterization of mechanisms as 'entities and activities organized such that they are productive of regular changes'. I showed that MDC demand, as per transparency, that all steps of a mechanism are causally connected (a demand for what they call 'productive continuity'). Machamer, Darden, and Craver, I showed, claim that we can give evidence for the productive continuity of mechanisms in molecular and neurobiology in terms of *regularities*. Knowing a link between two steps of the causal chain occurs regularly (that is, as per MDC's definition of 'regularly', it occurs in more than one case) is MDC's preferred source of evidence that that link is genuinely causal. I argued that while this may well be a source of evidence for productive continuity in molecular and neurobiology, it is by no means the only source of such evidence. In fact, the productive continuity of social mechanisms is hardly ever supported by such strict regularities as underlie molecular and neurobiology. I illustrated this by means of the case study of the Fashoda crisis in Schultz's work on coercive diplomacy, in which the productive continuity of the mechanism 'signalling' is established by means of smoking gun tests that rely on the close timings of the events of the crisis, and not by means of the recurrence of links in the chain of events in other cases. The only regularity in the Fashoda crisis is the mechanism *itself*, which Schultz argues recurs in other conflicts, and not the *links* between the events of the Fashoda crisis, which are arguably unique to the Fashoda case.

The main thread running through both these chapters was thus the question of how one can give evidence for transparency. Chapter 3 made clear how interventionists would give evidence for transparency, i.e. in terms of interventions, an idea which is then taken on in chapter 6 to support an argument against another framework of evidence for transparency, i.e. evidence in terms of regularities. A key question to move forward the discussion of both chapters is how to analyse the *comparability* of cases; in chapter 3, any evidence for a real, natural, or hypothetical intervention

relies on a comparison between two or more case studies, and in chapter 6, knowing to what extent links in a causal chain are comparable makes or breaks any reliance on regularities. This brings me to the second aspect of good research design that I set out to investigate: comparability.

2. Comparability

In the introduction of this thesis I discussed two aspects of comparability: *descriptive* comparability, i.e. the demand that the cause and effect in the cases in our study sample must refer to the same kinds of things, and *causal* comparability, i.e. the demand that all causal relations between cause and effect are the same (e.g. equally strong) in the cases in our study sample. Here, I will analyse both whether these demands are met by current best practice in process tracing, and how these demands are connected to the demand for transparency.

2.1 Descriptive comparability

My conclusions in chapter 2 shed further light on the demand for descriptive comparability. Setting out the limits to descriptive comparability, I argued that social science phenomena are both ambiguous (i.e. they can be systematized in different, opposing ways) and fuzzy (i.e. even once systematized with a particular concept, they do not have straightforward necessary and sufficient boundary conditions). I argued that systematizing a concept depends on the context and aims of our research. Therefore, one might argue that to demand that causes and effects refer to the same kinds of things in different cases is too strong. Calling a conflict in two different countries a 'civil war' does not imply that the conflicts are exactly the same.

These 'pluralist' limitations on descriptive comparability were further highlighted in my example of Collier and Hoeffler's definition of civil war, a systematization of the concept adopted from a study by Kristian Gleditsch, which in turn was based on

a systematization by Meredith Sarkees. However, I also used the example to illustrate that pluralism of systematization is constrained: which systematization one allows for a phenomenon like 'civil war' depends on the causal structure of civil war cases. If two conflicts we both call 'civil wars' turn out to be *too* different (e.g. have different causes), we ought to make sure that our systematized concepts respect that difference: it may mean that, in that study's context, we distinguish two subtypes of civil wars.

2.2 Causal comparability

Chapters 4 and 5 shed further light on causal comparability, the demand that all relations between cause and effect are the same in our sample. I discussed this demand in terms of generalization: under what circumstances we are allowed to generalize over a set of cases depends, after all, on how we compare the causal relations in those cases. General causal claims, I argued, assert that a relation holds between event types. I distinguished two different interpretations of generalization, the stronger Hitchcock causal homogeneity condition and the weaker Hausman average effect condition. The Hitchcock causal homogeneity condition makes individual-level claims, effectively stating that all individual cases within the domain are similar in their response to the cause variable of interest. The Hausman average effect condition, on the other hand, makes the domain-level claim that on average, the probability of the effect given the cause is larger than the probability of the effect in the absence of the cause. I showed that while it is difficult to establish a Hitchcockian generalization in the social sciences, these generalizations are useful in decision making; and while it is much easier to establish a Hausmanian generalization, these generalizations are less useful in decision making.

In Chapter 4, I asked in which situations we have evidence for Hitchcockian and Hausmanian generalizations. In chapter 5, I provided a synthesis of the two interpretations by making a distinction between ontological causal homogeneity and epistemic causal homogeneity. A set is ontologically causally homogeneous if

that set meets the Hitchcock causal homogeneity condition. Wesley Salmon, in response to the reference class problem, has taught us that if we know that a set of cases is not ontologically causally homogeneous, we will know a way of partitioning the set into subsets, each of which responds differently to some cause of interest. Following Salmon, I therefore characterized having incomplete information about whether a set meets the Hitchcock causal homogeneity condition in terms of not knowing a relevant partitioning for that set. I called such a set epistemically causally homogeneous. Thus, whether a set is epistemically causally homogeneous depends on our knowledge of that set.

I argued, in chapter 5, that there are several aspects to epistemic causal homogeneity. One can delineate a class of events or phenomena according to one property ('Is this class of conflicts homogeneous with reference to ethnic diversity?') or according to all properties ('Is there a class of civil wars homogeneous with reference to all causally relevant variables?'). This I called the difference between *relative* epistemic homogeneity, and *total* epistemic homogeneity. A second aspect to epistemic homogeneity is the distinction between *pragmatic* epistemic homogeneity, which takes into account the practical constraints on how much time and effort one can put into treating a heterogeneous set's subsets differently, and *true* epistemic causal homogeneity, which refers only to the case when a researcher genuinely does not know a relevant partitioning.

2.3 Comparability and process tracing

Do the current standards for best practice in process tracing meet the demands of descriptive and causal comparability? Yes and no.

As I have argued, the pluralist position in philosophy of measurement may lead us to conclude that there are different and equally valid ways one could systematize a particular social science phenomenon, depending on the context and aims of one's study; moreover, some social science phenomena do not have clear-cut boundaries and thus do not lend themselves to (conceptual) analysis in terms of necessary and sufficient conditions. To that extent, descriptive comparability is impossible unless

one assumes very clearly from the start one particular set of necessary and sufficient conditions, knowing full well that any appropriate systematization of the phenomenon is a fuzzy one. Chapters 4 and 5 teach us that even if we ignore the difficulties surrounding the systematization of cause and effect terms in the social sciences, most likely there will be causally relevant differences between the cases we study, i.e. the relations between cause and effect will be different in each case.

These general conclusions plausibly hold for all causal inquiry in the social sciences, not just for process tracing. What, then, can we say to process tracers to help them meet these challenges? In chapter 3, I defined process tracing as tracing the observable implications of a causal chain connecting a putative cause and effect of interest, which itself is the result of a (combination of different) underlying causal mechanism(s). Comparability is important for process tracing in two respects: firstly, process tracers who wish to generalize over more than one case should find evidence to support that the observable implications in all cases are implications of the same variables. If, for instance, the building of a training camp in Chechnya is a 'trace' of a very different variable than the building of a training camp in Sierra Leone, then it is not sensible to draw comparisons between the two despite this superficial commonality. This is the demand for descriptive comparability applied to process tracing.

Recall that whereas research designs which satisfy causal comparability will (by definition) satisfy descriptive comparability, research designs which satisfy descriptive comparability do not necessarily satisfy causal comparability. *Causal* comparability is important for process tracing in two respects. Firstly, causal comparability is relevant to any generalization from one case study to a set of target cases. We have seen examples of this in chapter 4, where I discussed Wood's hypothesis about the effects of defiance in a number of cases other than her test case, El Salvador. The second respect in which causal comparability is important for process tracing is for *transparency*: we need comparability to be able to draw some conclusion about whether the links of a causal chain are genuinely causal. Let us consider one final (very brief) example to make the link between transparency and

comparability concrete, and to show the distinction between what I shall call 'comparability for generalization' and 'comparability for transparency'.

2.3.1 The nuclear taboo

In her 1999 article "The Nuclear Taboo: The United States and the Normative Basis of Nuclear Non-Use", Nina Tannenwald presents an explanation for the Americans' non-use of nuclear weapons during the Korean War, the Vietnam War, and the Persian Gulf War. She argues that they refrained because the use of nuclear weapons had become constrained by a so-called 'nuclear taboo' after their use of such weapons at the end of the Second World War, i.e. it had become constrained by people's revulsion against the use of nuclear weapons (see also Bennett and Checkel 2015, Collier 2011a, Tannenwald 1999). Tannenwald uses the American case study to test a more general hypothesis, viz. that a nuclear taboo is a stronger explanation for a state's non-use of nuclear weapons than alternative explanations like deterrence (i.e. threats by one party aimed at dissuading another party from using nuclear weapons). The causal mechanisms behind the non-use of nuclear weapons are, Tannenwald argues, "domestic public opinion, [adverse] world opinion (…), and personal conviction informed by beliefs about American values and conceptions of the appropriate behavior of civilized nations" (Tannenwald 1999, 462).

In brief, Tannenwald traces the following observable implications of these causal mechanisms. During the Korean War the 'nuclear taboo' first manifested itself in the personal "moral concerns" that President Truman and his advisors had about using such a "disproportionate" weapon (Tannenwald 1999, 446). Later, such moral concerns were taken up by the public, and public revulsion in turn constrained President Eisenhower and his advisors, since using weapons in spite of public aversion would have a high political cost. Tannenwald describes how, over time, the taboo became entrenched in both domestic public opinion and world opinion,

so that during the later Vietnam War and Persian Gulf War, American leaders did not even come close to using nuclear weapons.

Comparability plays a dual role in Tannenwald's work; as such, her study provides an illustration of my claim that comparability has two implications for process tracing. Firstly, comparability is an important factor for generalization: in her conclusion, Tannenwald speculates on the generalizability of her claims beyond the US case. To do so, she draws upon similarities between the US and other states: "Because the United States is an open democracy, penetrated by domestic opinion and ideas, and with a perceived tradition of humanitarian rights and values, it may in this sense be an 'easier' case (…). This suggests that if the taboo operates in the United States, it probably operates in other democracies less committed to, and reliant on, nuclear weapons historically." (Tannenwald 1999, 464) To give evidence of the causal role of the nuclear taboo in other democratic states, Tannenwald argues, we ought to give evidence of the similarity of these states. We can therein ignore some of the causal differences between these states; it does not matter how much less they are committed to the use of nuclear weapons, since they simply have to meet the threshold of *being less committed than the US.* On the other hand, whether a state is a democracy *is* relevant to the operating of the causal mechanisms: Tannenwald speculates that public opinion will have a different impact on the use or non-use of nuclear weapons in non-democratic states. Arguably, non-democratic governments will be less concerned about the repercussions of going against public opinion.

Comparability also plays another role in Tannenwald's study, namely in the transparency of her analysis of the US case study. For instance, if it were to turn out that there is no difference in the motivations for non-use of nuclear weapons in non-democratic states, this would cast doubt on Tannenwald's mechanism in the detailed case study of the United States, namely that it depends on public opinion within a democracy. This latter aspect of Tannenwald's study is just one illustration of the importance of comparability for transparency. I have shown in several parts of this thesis that knowledge of the similarities and differences between cases, i.e. knowledge of descriptive and causal comparability, helps us to better understand

whether the process we postulate is genuinely causal. Other examples include Bakke's case study in chapter 3 (where it turned out to be helpful if we could compare the Chechen insurgents' attitudes towards radical tactics before and after the arrival of Islamist insurgents).

So, to sum up, comparability is important for those process tracers who aim to directly infer general causal claims. Comparability is also important more generally for process tracing, insofar as it provides support for transparency. I argued that we need comparability to be able to draw some conclusion about whether the links of a causal chain are genuinely causal. Comparability for transparency need only concern the case at hand and a relevant counterfactual case. This is, in a sense, a limited comparability requirement. Think back to the Bakke case study: all one needs to worry about in this case is what would happen if the transnational insurgents had not used relational diffusion or brokerage. To find out about what would happen, we need a comparable case; however, as we have seen in chapter 3, this case can often be a hypothetical one, and as such can be as similar as a researcher requires. (And, as such, we do not require anything as strict as MDC's 'regularities', discussed in chapter 6.) When it comes to generalization, on the other hand, one needs to know more, i.e. the comparability between the case under study and some target case(s).

3. Recommendations for process tracing

Let me now use the conclusions above to provide two constructive recommendations for process tracers in practice.

3.1 Recommendation 1

To give my first concrete recommendation, I wish to refer to the most recent list of recommendations for process tracers, Bennett and Checkel's ten recommendations

for process tracing[1]. Two of Bennett and Checkel's recommendations are directly relevant to my conclusions regarding transparency and comparability. Firstly, Bennett and Checkel argue that researchers ought to "[u]se deduction to ask 'if my explanation is true, what will be the specific process leading to the outcome?'" (Bennett and Checkel 2015, 30). Secondly, they argue that researchers ought to "[c]ombine process tracing with case comparisons when useful for the research goal and feasible" (Bennett and Checkel 2015, 29).

The first recommendation, to study the deductive implications of a mechanism in a case study, was part of my initial formalization of process tracing in chapter 3. Let me therefore immediately turn to Bennett and Checkel's second recommendation that one ought to combine process tracing with case comparisons. This recommendation is closely related to what I discussed as 'comparability for transparency' above. To illustrate Bennett and Checkel's recommendation, think back to the Tannenwald study. In their chapter, Bennett and Checkel argue that we can use case comparisons for process tracing by comparing two 'most-similar' cases. In the Tannenwald case, an example would be a study of one democratic state, and one non-democratic state, both of which are 'similar' in all other respects. If there is a nuclear taboo in the democratic state, but not the other, this provides evidence for the claim that democracy is a difference-maker for the operation of Tannenwald's public opinion mechanism that leads to the non-use of nuclear weapons.

Yet, as we have seen in chapters 3 and 6, there are other ways of collecting evidence for the genuinely causal nature of the links of the chain under study. Combining, then, the first and second recommendation in Bennett and Checkel with my own analyses in this thesis, I would refine Bennett and Checkel's recommendation as follows: *Combine the study of observable implications of a mechanism with any available evidence that the links of this mechanism each express a genuine causal relationship, whether this evidence is derived from case comparisons or otherwise.*

---

[1] Bennett and Checkel spend one paragraph on most of these considerations. We should therefore not overstate their importance.

3.2 Recommendation 2

Bennett and Checkel, in their ten recommendations for best practice, do not provide any particular recommendation that answers to my concerns regarding comparability for generalization. For that purpose, I would add the following recommendation to their list: *When generalizing beyond a single case study to a more general set of cases, take into account any knowledge you may have of causally relevant differences within that set.* The further sophistication of this recommendation would then follow my taxonomy of epistemic homogeneity as introduced in chapter 4 and outlined in chapter 5.

I would spell out this recommendation as follows. Like all social scientists, process tracers should look for descriptive and causal comparability whilst respecting the epistemic limitations they are working under. Epistemic homogeneity, then, may be a more workable requirement for research design than the causal comparability demand as defined by John Gerring, i.e. more workable than the demand that we choose "cases that are similar to each other in whatever ways might affect the *Y* or the posited $X \rightarrow Y$ relationship" (Gerring 2005, 185).

4. Suggestions for further research

To end this conclusion, I wish to point out two considerations for further research which build upon, and go beyond, the arguments I have presented in this thesis. Firstly, as I indicated in the introduction, the work in this thesis is centred mainly on one area of the social sciences, conflict studies. My choice for this area of the social sciences was inspired by two considerations: firstly, the wish to provide some consistency across chapters, and secondly, the fact that process tracing is often used by conflict scholars, as well as described by methodologists who were trained in conflict studies. Yet, process tracing is also used in other areas of the social sciences, from educational research to sociology, and it is therefore of some interest to study if, and how, the method differs in those areas. Moreover, it is of considerable interest to ask whether the conclusions in this thesis, e.g. as they relate to

transparency and comparability, are applicable outside conflict studies. I have given some indication as to why I believe the conclusions apply more generally: for instance, the claim that phenomena are both 'ambiguous' and 'fuzzy' extends beyond conflict studies. It would nevertheless be worthwhile to see how these issues constrain other social scientific research in practice.

Secondly, in parts of this thesis I have focused on comparability across cases, but I have not focused on what some consider a closely related aspect of causal inquiry: *abstraction*. Abstraction is discussed amongst others by Nancy Cartwright and Jeremy Hardie, who argue that whether one is able to compare across cases depends on whether one has characterized each case at the right level of abstraction. Consider, again, the example from chapter 3, Kirstin Bakke's study of radicalization. While, perhaps, the causal link 'watching suicide terrorism videos causes an increase in suicide terrorism' applies in the Chechen case only, it may be that a description of this causal link at a higher level of abstraction (e.g. as 'watching terrorism videos causes an increase in terrorism') applies to a wider set of conflicts. Cartwright and Hardie thus argue that the scope conditions of a causal claim will depend largely on "finding concepts at the right level of abstraction, or generality" (Cartwright and Hardie 2012, 79); they call this search for the right level of abstraction 'vertical search'.

In this thesis, I have shown that process tracers ought to have comparability on their mind even if they just want to explain the causal chain in a particular case study, i.e. that comparability matters for the transparency of a single case study. Arguably, good explanation of a particular case is not just about describing the actual causal process, but also about using the right level of abstraction for this description. It would be of interest to expand on how the demand for comparability relates to abstraction and Cartwright's notion of 'vertical search'. If indeed abstraction is an important element of what counts as good causal explanation of a case, vertical search ought to be of interest to process tracers. However, the relation between abstraction and mechanisms ought to be investigated further before one can use vertical search in process tracing.

References

Adcock, Robert, and David Collier (2001) "Measurement Validity: a Shared Standard for Qualitative and Quantitative Research." *The American Political Science Review* 95(3): 529-546.

Alexandrova, Anna (2008) "First-Person Reports and the Measurement of Happiness." *Philosophical Psychology* 21(5): 571-583.

Andersen, Holly (2014a) "A Field Guide to Mechanisms: Part I." *Philosophy Compass* 9(4): 274-283.

Andersen, Holly (2014b) "A Field Guide to Mechanisms: Part II." *Philosophy Compass* 9(4): 284-293.

Ångström, Jan (2001) "Towards a Typology of Internal Armed Conflict: Synthesising a Decade of Conceptual Turmoil." *Civil Wars* 4(3): 93-116.

Bakke, Kristin M. (2013) "Copying and Learning from Outsiders? Assessing Diffusion from Transnational Insurgents in the Chechen Wars." In *Transnational Dynamics of Civil War*, edited by Jeffrey T. Checkel. 31-62. Cambridge: Cambridge University Press.

Bechtel, William, and Adele Abrahamsen (2005) "Explanation: a Mechanist Alternative." *Studies in History and Philosophy of Biological and Biomedical Sciences* 36(2): 421-441.

Bechtel, William, and Robert C. Richardson (1993) *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton: Princeton University Press.

Bennett, Andrew (2010) "Process Tracing and Causal Inference." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited by Henry E. Brady and David Collier, 2nd ed. 207-220. Lanham: Rowman & Littlefield.

Bennett, Andrew, and Jeffrey T. Checkel (2015) "Process Tracing: from Philosophical Roots to Best Practices." In *Process Tracing: from Metaphor to Analytic Tool*, edited by Andrew Bennett and Jeffrey T. Checkel. 3-38. Cambridge: Cambridge University Press.

Bennett, Andrew, and Colin Elman (2006) "Qualitative Research: Recent Developments in Case Study Methods." *Annual Review of Political Science* 9: 455-476.

Bennett, Andrew, and Colin Elman (2007) "Case Study Methods in the International Relations Subfield." *Comparative Political Studies* 40(2): 170-195.

Bennett, Andrew, and Alexander L. George (1997) "Process Tracing in Case Study Research." *MacArthur Foundation Workshop on Case Study Methods*, Belfer Center for Science and International Affairs, Harvard University, October 17-19 1997.

Bogen, James (2005) "Regularities and Causality; Generalizations and Causal Explanations." *Studies in History and Philosophy of Biological and Biomedical Sciences* 36(2): 397-420.

Bogen, James (2008) "Causally Productive Activities." *Studies in History and Philosophy of Science* 39(1): 112-123.

Brady, Henry E. (2008) "Causation and Explanation in Social Science." In *The Oxford Handbook of Political Methodology*, edited by Janet M. Box-Steffensmeier, Henry E. Brady and David Collier. 217-270. Oxford: Oxford University Press.

Brady, Henry E., and David Collier (2004) *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. 1st ed. Lanham: Rowman & Littlefield.

Brady, Henry E., and David Collier (2010) *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. 2nd ed. Lanham: Rowman & Littlefield.

Bulmer, Martin (2001) "Social Measurement: What Stands in Its Way?" *Social Research* 68(2): 455-480.

Bunge, Mario (1997) "Mechanism and Explanation." *Philosophy of the Social Sciences* 27(4): 410-465.

Cartwright, Nancy (1979) "Causal Laws and Effective Strategies." *Noûs* 13(4): 419-437.

Cartwright, Nancy (1989) *Nature's Capacities and their Measurement*. Oxford: Clarendon Press.

Cartwright, Nancy (2001) "Modularity: It Can - and Generally Does - Fail." In *Stochastic Causality*, edited by Maria C. Galavotti, Patrick Suppes and Domenico Constantini. 65-84. Stanford: CSLI Publications.

Cartwright, Nancy (2002) *Causation: One Word; Many Things*. London: Centre for Philosophy of Natural and Social Science. Technical Report 07/03.

Cartwright, Nancy (2007) *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge: Cambridge University Press.

Cartwright, Nancy, and Norman M. Bradburn (ms.) "A Theory of Measurement." Accessed October 2 2014 at dur.ac.uk/resources/philosophy/BradburnCartwrightThMeaspostSSFINAL.pdf.

Cartwright, Nancy, and Sophia Efstathiou (2011) "Hunting Causes and Using Them: Is There No Bridge from Here to There?" *International Studies in the Philosophy of Science* 25(3): 223-241.

Cartwright, Nancy, and Jeremy Hardie (2012) *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford: Oxford University Press.

Cartwright, Nancy, and Rosa W. Runhardt (2014) "Measurement." In *Philosophy of Social Science: a New Introduction*, edited by Nancy Cartwright and Eleonora Montuschi. 265-287. Oxford: Oxford University Press.

Chamberlain, Muriel E. (2013) "Fashoda and the Anglo-French Agreements of 1904." In *The Scramble for Africa*, 3rd ed. 79-84. Oxon: Routledge.

Checkel, Jeffrey T. (2008) "Process Tracing." In *Qualitative Methods in International Relations: A Pluralist Guide*, edited by Audie Klotz and Deepa Prakash. 114-127. New York: Palgrave Macmillan.

Checkel, Jeffrey T. (2013) "Transnational Dynamics of Civil War." In *Transnational Dynamics of Civil War*, edited by Jeffrey T. Checkel. 3-27. Cambridge: Cambridge University Press.

Clarke, Brendan (2011) *Causality in Medicine with Particular Reference to the Viral Causation of Cancers*. PhD Thesis, Department of Science and Technology Studies, University College London, London.

Collier, David (2011a) "Teaching Process Tracing: Examples and Exercises." *PS: Political Science and Politics* 44(4).

Collier, David (2011b) "Understanding Process Tracing." *PS: Political Science and Politics* 44(4): 823-830.

Collier, Paul, and Anke Hoeffler (2000) *Greed and Grievance in Civil War*. World Bank Policy Research. Working Paper 2355.

Collier, Paul, and Anke Hoeffler (2001) "Data Issues in the Study of Conflict." *Data Collection on Armed Conflict*, Uppsala, 8-9 June 2001.

Collier, Paul, and Anke Hoeffler (2004) "Greed and Grievance in Civil War." *Oxford Economic Papers* 56(4): 563-595.

Collier, Paul, Anke Hoeffler, and Dominic Rohner (2009) "Beyond Greed and Grievance: Feasibility and Civil War." *Oxford Economic Papers* 61(1): 1-27.

Collier, Paul, and Nicholas Sambanis, eds. (2005) *Understanding Civil War: Evidence and Analysis*. Vol. 1. Africa. Washington: The World Bank.

Cramer, Christopher (2006) *Civil War is Not a Stupid Thing: Accounting for Violence in Developing Countries*. London: Hurst & Co.

Crasnow, Sharon (2012) "The Role of Case Study Research in Political Science: Evidence for Causal Claims." *Philosophy of Science* 79(5): 655-666.

Crasnow, Sharon (forthcoming) "The Measure of Democracy: Coding in Political Science." In *Standardization in Measurement: Philosophical, Historical and Sociological Issues*, edited by Lara Huber and Oliver Schlaudt. London: Pickering & Chatto Publishers.

Darden, Lindley (2008) "Thinking Again about Biological Mechanisms." *Philosophy of Science* 75(5): 958-969.

Darden, Lindley, and Carl F. Craver (2002) "Strategies in the Interfield Discovery of the Mechanism of Protein Synthesis." *Studies in History and Philosophy of Biological and Biomedical Sciences* 33(1): 1-28.

Dupré, John (1984) "Probabilistic Causality Emancipated." *Midwest Studies in Philosophy* 9(1): 169-175.

Dupré, John (1990) "Probabilistic Causality: A Rejoinder to Ellery Eells." *Philosophy of Science* 57(4): 690-698.

Dupré, John, and James Woodward (2013) "Mechanism and Causation in Biology." *Proceedings of the Aristotelian Society* Supplementary Volume LXXXVII: 19-37.

Eckstein, Harry (1964) "Introduction: Toward the Theoretical Study of Internal War." In *Internal War: Problems and Approaches*. 1-32. New York: The Free Press of Glencoe.

Eckstein, Harry (1965) "On the Etiology of Internal Wars." *History and Theory* 4(2): 133-163.

Eells, Ellery (1991) *Probabilistic Causality*. Cambridge: Cambridge University Press.

Fearon, James D. (1994) "Domestic Political Audiences and the Escalation of International Disputes." *The American Political Science Review* 88(3): 577-592.

Fearon, James D. (2005) "Primary Commodity Exports and Civil War." *The Journal of Conflict Resolution* 49(4): 483-507.

Fearon, James D., and David D. Laitin (2011) "Integrating Qualitative and Quantitative Methods." In *The Oxford Handbook of Political Science*, edited by Robert E. Goodin. 1166-1186. Oxford: Oxford University Press.

Gallie, Walter B. (1956) "Essentially Contested Concepts." *Proceedings of the Aristotelian Society* 56: 167-198.

George, Alexander L. (1979) "Case Studies and Theory Development: The Method of Structured, Focused Comparison." In *Diplomacy: New Approaches in History, Theory, and Policy*, edited by Paul G. Lauren. 43-68. New York: Free Press.

George, Alexander L., and Andrew Bennett (2005) *Case Studies and Theory Development in the Social Sciences*. Cambridge: MIT Press.

Gerring, John (2005) "Causation: A Unified Framework for the Social Sciences." *Journal of Theoretical Politics* 17(2): 163-198.

Gerring, John (2008a) "Book Review: Rethinking Social Inquiry: Diverse Tools, Shared Standards." *The Journal of Politics* 70(1): 274-276.

Gerring, John (2008b) "Review Article: The Mechanismic Worldview: Thinking Inside the Box." *British Journal of Political Science* 38(1): 167-179.

Gerring, John, and Craig Thomas (2011) "Quantitative and Qualitative: a Question of Comparability." In *International Encyclopedia of Political Science*, edited by

Bertrand Badie, Dirk Berg-Schlosser and Leonardo Morlino. London: SAGE Publications.

Gleditsch, Kristian S. (2004) "A Revised List of Wars Between and Within Independent States, 1816-2002." *International Interactions* 30(3): 231-262.

Gleditsch, Nils P., Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand (2002) "Armed Conflict 1946-2001: A New Dataset." *Journal of Peace Research* 39(5): 615-637.

Glennan, Stuart S. (1996) "Mechanisms and the Nature of Causation." *Erkenntnis* 44: 49-71.

Glennan, Stuart S. (2002) "Rethinking Mechanistic Explanation." *Philosophy of Science* 69(S3): S342-S353.

Guala, Francesco (2010) "Extrapolation, Analogy, and Comparative Process Tracing." *Philosophy of Science* 77(5): 1070-1082.

Gurr, Ted R., and Will H. Moore (1997) "Ethnopolitical Rebellion: A Cross-Sectional Analysis of the 1980s with Risk Assessments for the 1990s." *American Journal of Political Science* 41(4): 1079-1103.

Hall, Peter A. (2013) "Symposium: Tracing the Progress of Process Tracing." *European Political Science* 12: 20-30.

Hausman, Daniel M. (2010) "Probabilistic Causality and Causal Generalizations." In *The Place of Probability in Science*, edited by Ellery Eells and J.H. Fetzer. 47-63. Dordrecht: Springer.

Hedström, Peter, and Richard Swedberg (1998) *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press.

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

Hesslow, Germund (1976) "Two Notes on the Probabilistic Approach to Causality." *Philosophy of Science* 43(2): 290-292.

Hitchcock, Christopher R. (1993) "A Generalized Probabilistic Theory of Causal Relevance." *Synthese* 97(3): 335-364.

Hitchcock, Christopher R. (1995) "The Mishap at Reichenbach Fall: Singular vs. General Causation." *Philosophical Studies* 78(3): 257-291.

Hitchcock, Christopher R. (2001) "A Tale of Two Effects." *The Philosophical Review* 110(3): 361-396.

Holland, Paul W. (1986) "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396): 945-960.

Horowitz, Donald L. (1985) *Ethnic Groups in Conflict*. Berkeley: University of California Press.

Horowitz, Shale, and Uk Heo (2008) "Do Ethnic and Non-Ethnic Civil Wars Have Distinct Causes?" *Annual Meeting of the International Studies Association*, San Francisco.

Hume, David (1748) *An Enquiry Concerning Human Understanding*. London: A. Millar.

Humphreys, Paul (1986) "Causality in the Social Sciences: An Overview." *Synthese* 68(1): 1-12.

Institute for Economics and Peace (2012) *Global Peace Index: 2012*. Accessed 23 August 2012 at economicsandpeace.org/wp-content/uploads/2011/09/2012-Global-Peace-Index-Report.pdf.

Jackson, Patrick T. (2011) *The Conduct of Inquiry in International Relations: Philosophy of Science and Its Implications for the Study of World Politics*. London: Routledge.

Johnson, James (2006) "Consequences of Positivism: A Pragmatist Assessment." *Comparative Political Studies* 39(2): 224-252.

Kincaid, Harold (2009) "Causation in the Social Sciences." In *The Oxford Handbook of Causation*, edited by Helen Beebee, Christopher Hitchcock and Peter Menzies. 726-743. Oxford: Oxford University Press.

King, Gary, Robert O. Keohane, and Sidney Verba (1994) *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.

Lacina, Bethany, and Nils P. Gleditsch (2005) "Monitoring Trends in Global Combat: a New Dataset of Battle Deaths." *European Journal of Population* 21: 145-166.

Levy, Arnon (2013) "Three Kinds of New Mechanism." *Biology and Philosophy* 28(1): 99-114.

Little, Daniel (1991) *Varieties of Social Explanation: An Introduction to the Philosophy of Social Science*. Boulder: Westview Press.

Little, Daniel (1993) "On the Scope and Limits of Generalizations in the Social Sciences." *Synthese* 97(2): 183-207.

Little, Daniel (2011) "Causal Mechanisms in the Social Realm." In *Causality in the Sciences*, edited by Phyllis McKay Illari, Federica Russo and Jon Williamson. 273-295. Oxford: Oxford University Press.

Luce, R. Duncan, and Patrick Suppes (2002) "Representational Measurement Theory." In *Stevens' Handbook of Experimental Psychology*, edited by J. Wixted and H. Pashler, 3rd ed. 1-41. New York: Wiley.

Lyall, Jason (2015) "Process Tracing, Causal Inference, and Civil War." In *Process Tracing: From Metaphor to Analytic Tool*, edited by Andrew Bennett and Jeffrey Checkel. 186-207. Cambridge: Cambridge University Press.

Machamer, Peter (2004) "Activities and Causation: the Metaphysics and Epistemology of Mechanisms." *International Studies in the Philosophy of Science* 18(1): 27-39.

Machamer, Peter, Lindley Darden, and Carl F. Craver (2000) "Thinking about Mechanisms." *Philosophy of Science* 67(1): 1-25.

Mackie, John L. (1974) *The Cement of the Universe: A Study of Causation*. Oxford: Clarendon Press.

Mahoney, James (2001) "Beyond Correlational Analysis: Recent Innovations in Theory and Method." *Sociological Forum* 16(3): 575-593.

McAdam, Doug, Sidney Tarrow, and Charles Tilly (2001) *Dynamics of Contention*. Cambridge: Cambridge University Press.

McKeown, Timothy J. (1999) "Case Studies and the Statistical Worldview: Review of King, Keohane, and Verba's Designing Social Inquiry: Scientific Inference in Qualitative Research." *International Organization* 53(1): 161-190.

Miller, Richard W. (1978) "Methodological Individualism and Social Explanation." *Philosophy of Science* 45(3): 387-414.

Morgan, Mary S. (2013) "Nature's Experiments and Natural Experiments in the Social Sciences." *Philosophy of the Social Sciences* 43(3): 341-357.

Murray, C.J.L., G. King, A.D. Lopez, N. Tomijma, and E.G. Krug (2002) "Armed Conflict as a Public Health Problem." *British Medical Journal* 324: 346-349.

Nicholson, Daniel J. (2012) "The Concept of Mechanism in Biology." *Studies in History and Philosophy of Biological and Biomedical Sciences* 43(1): 152-163.

Northcott, Robert (2008) "Causation and Contrast Classes." *Philosophical Studies* 139(1): 111-123.

Pearl, Judea (2000) *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Psillos, Stathis (2004) "A Glimpse of the Secret Connexion: Harmonizing Mechanisms with Counterfactuals." *Perspectives on Science* 12(3): 288-319.

Ragin, Charles C. (2008) *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: University of Chicago Press.

Reiss, Julian (2005) "Causal Instrumental Variables and Interventions." *Philosophy of Science* 72: 964-976.

Reiss, Julian (2007) "Do We Need Mechanisms in the Social Sciences?" *Philosophy of the Social Sciences* 37(2): 163-184.

Reiss, Julian (2008a) *Error in Economics: Towards a More Evidence-Based Methodology*. London: Routledge.

Reiss, Julian (2008b) "Social Capacities." In *Nancy Cartwright's Philosophy of Science*, edited by Luc Bovens, Carl Hoefer and Stephen Hartmann. 265-288. London: Routledge.

Reiss, Julian (2009a) "Causation in the Social Sciences: Evidence, Inference, and Purpose." *Philosophy of the Social Sciences* 39(1): 20-40.

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

Runhardt, Rosa W. (forthcoming) "Evidence for Causal Mechanisms in Social Science: Recommendations from Woodward's Manipulability Theory of Causation." *Philosophy of Science*.

Salmon, Wesley C. (1971) *Statistical Explanation and Statistical Relevance*. Pittsburgh: University of Pittsburgh Press.

Sambanis, Nicholas (2001) "Do Ethnic and Nonethnic Civil Wars Have the Same
     Causes? A Theoretical and Empirical Enquiry (Part I)." *The Journal of Conflict
     Resolution* 45(3): 259-282.

Sambanis, Nicholas (2004a) *Civil War Coding Notes*. Accessed August 14 2012 at
     pantheon.yale.edu/~ns237/index/research/CivilWarCodingNotes.pdf.

Sambanis, Nicholas (2004b) "What is Civil War?" *The Journal of Conflict Resolution*
     48(6): 814-858.

Sarkees, Meredith Reid, and Phil Schafer (2000) "The Correlates of War Data on
     War: An Update to 1997." *Conflict Management and Peace Science* 18(1): 123-
     144.

Schultz, Kenneth A. (2001) *Democracy and Coercive Diplomacy*. Cambridge:
     Cambridge University Press.

Searle, John R. (2006) "Social Ontology: Some Basic Principles." *Anthropological
     Theory* 6(1): 12-29.

Searle, John R. (2010) *Making the Social World: the Structure of Human Civilization*.
     Oxford: Oxford University Press.

Searle, John R. (2011) "Book Symposium, Making the Social World: the Structure of
     Human Civilization." *Analysis* 71(4): 695-696.

Spirtes, Peter, Clark Glymour, and Richard Scheines (1993) *Causation, Prediction, and
     Search*. New York: Springer-Verlag.

Spirtes, Peter, and Richard Scheines (2003) *Causal Inference of Ambiguous
     Manipulations*. Pittsburgh: Department of Philosophy, Carnegie Mellon
     University. Technical Report 138.

Spirtes, Peter, and Richard Scheines (2004) "Causal Inference of Ambiguous
     Manipulations." *Philosophy of Science* 71(5): 833-845.

Steel, Daniel (2004) "Social Mechanisms and Causal Inference." *Philosophy of the
     Social Sciences* 34(1): 55-78.

Steel, Daniel (2008) *Across the Boundaries: Extrapolation in Biology and Social Science*.
     Oxford: Oxford University Press.

Steele, Katie, Yohay Carmel, Jean Cross, and Chris Wilcox (2009) "Uses and Misuses of Multicriteria Decision Analysis (MCDA) in Environmental Decision Making." *Risk Analysis* 29(1): 26-33.

Stevens, Stanley Smith (1946) "On the Theory of Scales of Measurement." *Science* 103(2684): 677-680.

Stevens, Stanley Smith (1951) "Mathematics, Measurement, and Psychophysics." In *Handbook of Experimental Psychology*. 1-49. New York: Wiley.

Suppes, Patrick (1998) "Theory of Measurement." In *Routledge Encyclopedia of Philosophy*, edited by Edward Craig. 243-249. London: Routledge.

Tal, Eran (2013) "Old and New Problems in Philosophy of Measurement." *Philosophy Compass* 8(12): 1159-1173.

Tannenwald, Nina (1999) "The Nuclear Taboo: The United States and the Normative Basis of Nuclear Non-Use." *International Organization* 53(3): 433-468.

Tilly, Charles (2001) "Mechanisms in Political Processes." *Annual Review of Political Science* 4: 21-41.

Vanhanen, Tatu (1999) "Domestic Ethnic Conflict and Ethnic Nepotism: A Comparative Analysis." *Journal of Peace Research* 36(1): 55-73.

Waldner, David (2012) "Process Tracing and Causal Mechanisms." In *The Oxford Handbook of Philosophy of Social Science*, edited by Harold Kincaid. 65-84. Oxford: Oxford University Press.

Watkins, John W. N. (1957) "Historical Explanation in the Social Sciences." *British Journal for the Philosophy of Science* 8(30): 104-117.

Wimmer, Andreas, Lars-Erik Cederman, and Brian Min (2009) "Ethnic Politics and Armed Conflict: a Configurational Analysis of a New Global Data Set." *American Sociological Review* 74(2): 316-337.

Wimsatt, William C. (1972) "Complexity and Organization." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1972: 67-86.

Wittgenstein, Ludwig (1953) *Philosophical Investigations*. Translated by G.E.M. Anscombe. Edited by Rush Rhees and G.E.M. Anscombe. Oxford: Blackwell.

Wold Bank (1990) *World Development Report: Poverty*. New York: Oxford University Press.

Wood, Elizabeth (2003) *Insurgent Collective Action and Civil War in El Salvador*. New York: Cambridge University Press.

Woodward, James (2002) "What Is a Mechanism? A Counterfactual Account." *Philosophy of Science* 69(S3): S366-S377.

Woodward, James (2003) *Making Things Happen: a Theory of Causal Explanation*. Oxford: Oxford University Press.

Woodward, James (2008) "Invariance, Modularity, and All That: Cartwright on Causation." In *Nancy Cartwright's Philosophy of Science*, edited by Luc Bovens, Carl Hoefer and Stephen Hartmann. 198-241. London: Routledge.

Woodward, James (forthcoming) "Methodology, Ontology, and Interventionism." *Synthese*: 1-23.

Ylikoski, Petri (2011) "Social Mechanisms and Explanatory Relevance." In *Analytical Sociology and Social Mechanisms*, edited by Pierre Demeulenaere. 154-172. Cambridge: Cambridge University Press.