Case selection techniques in Process-tracing and the implications of taking the study of causal mechanisms seriously

Derek Beach
Associate professor, PhD
Department of Political Science
University of Aarhus

Rasmus Brun Pedersen
Associate professor, PhD
Department of Political Science and Department of Business Communication
University of Aarhus
derek@ps.au.dk, brun@ps.au.dk

Abstract

This paper develops guidelines for each of the three variants of Process-tracing (PT): explaining outcome PT, theory-testing, and theory-building PT. Case selection strategies are not relevant when we are engaging in explaining outcome PT due to the broader conceptualization of outcomes that is a product of the different understandings of case study research (and science itself) underlying this variant of PT. Here we simply select historically important cases because they are for instance the First World War, not a 'case of' failed deterrence or crisis decision-making. Within the two theory-centric variants of PT, typical case selection strategies are most applicable. A typical case is one that is a member of the set of X, Y and the relevant scope conditions for the mechanism. We put forward that pathway cases, where scores on other causes are controlled for, are less relevant when we take the study of mechanisms seriously in PT, given that we are focusing our attention on how a mechanism contributes to produce Y, not on the causal effects of an X upon values of Y. We also discuss the role that deviant cases play in theory-building PT, suggesting that PT cannot stand alone, but needs to be complemented with comparative analysis of the deviant case with typical cases.

1. Introduction

Despite the widespread use and methodological writing about Process-tracing (PT), there has so far been little progression in the development of case selection strategies for PT. Our argument is that existing guidelines for selection strategies in the broader field of case study research are often inappropriate for two reasons. First, the existing literature does not distinguish clearly among the different variants of PT, thereby producing relatively murky prescriptions that do not always match the distinct research purposes of each variant. Second, if we take the study of mechanisms in PT seriously, existing guidelines that detail many different types of cases (deviant, pathway, etc) are less relevant. The point here is that existing case selection strategies are more in-line with research designs aimed at investigating causal effects and the difference that causal factors make on the outcome. In PT we are interested in the mechanisms linking X with Y. Therefore we argue that the most common selection strategy for PT should be choosing typical cases where both X and Y are present.

In comparison to other small-n comparative or single case methods, PT as a distinct research method involves, '…attempts to identify the intervening causal process – the causal chain and causal mechanism…' (George and Bennett 2005:206-207) that links a cause (or set of causes and an outcome. The promise of PT as a methodological tool is that it enables the researcher to study more-or-less directly the *causal mechanism* linking a causal condition (or set of conditions)[1] and an outcome, allowing us to open up the 'black box' of causality (Gerring 2007: 45) and focus on how X or a set of X's actually produces an outcome. However, if we take seriously the ambition of PT to trace causal mechanisms, this has serious implications for case selection strategies.

Other small-n methods such as a most-similar-systems design aim at assessing the causal effect a given X has on Y. For instance if we want to know *whether* economic development contributes to democracy (i.e. the causal effect of X on Y), we can adopt a comparative most-similar-systems design, comparing a case of low development with a case of high development, seeing if we find the hypothesized variation in Y while holding other factors constant. If we have reasons to expect that the chosen cases are typical cases, we can then make the inference that X is causally related to Y in the broader population of the phenomenon.

The point is that we would be investigating *whether* X contributes to *producing a difference*

1 - We use the term condition/outcome instead of independent/dependent variable to mark the fact that we are dealing with invariant designs, and to denote the difference in how conditions and variables are conceptualized (see Goertz, 2006 for more on the second point).

in the value of Y. If we want to know whether X produces a difference, we naturally have to have variation in the outcome. King, Keohane and Verba state correctly that, 'When observations are selected on the basis of a particular value of the dependent variable, nothing whatsoever can be learned about the causes without taking into account other values [of Y]' (King, Keohane and Verba, 1994: 129).

Yet PT is a tool to investigate *how* X contributes to produce Y through a causal mechanism or mechanisms.[2] We want to understand what it is about the arrow in-between X and Y that transmits causal forces to contribute to producing Y. If we are studying the mechanisms in-between X and Y, we already either know that there is a causal relationship between X and Y, or have good theoretical and/or empirical grounds for assuming that there is one. The goal of PT is to either deductively or inductively explore *how* X contributes to produce Y through the operation of a causal mechanism.

The distinction between studying *whether* and *how* can be illustrated using the classic example of the relationship between smoking (X) and cancer (Y). Studying *whether* smoking is causally related to cancer involves an experimental research design, where we investigate the impact that the treatment (X) and control (not X) have upon incidences of cancer in a large number of persons. Choosing only to investigate patients in a cancer ward (a positive on outcome design) would not enable us to detect whether X produces a difference in the outcome, unless we can assume that smoking is the only cause of cancer. We would want to have full variation in the outcome across a range of comparable cases to measure the mean causal effect of X upon the outcome, ideally by selecting a large number of randomly chosen cases to ensure variation on X and Y while controlling for potential confounders.

Investigating *how* smoking contributes to producing cancer is a very different research question. We take as a starting point research that tells us that X produces a difference in the value of Y, focusing our attention on how smoking contributes to causing cancer by deliberately selecting only smokers (X is present). Naturally we need to possess information about how 'healthy' lungs operate to make sense of the process whereby exposure to chemicals from cigarettes causes lung cell mutations that become cancerous cells; information that enables us to spot what wrong. But studying non-smokers will tell us little about the actual mechanism whereby smoking produces cancer. This is at the core of our argument that when engaging in PT, we usually will only select

2 Note we use the term 'contribute', as we are not assuming that a given X (and causal mechanism) is sufficient to produce Y. While theories of necessity are widespread in political science (see Goertz, 2003), theories of sufficient causal relations are rarer, although many applications of QCA analyze sufficiency. As regards broader theorization about sufficient relationships, the democratic peace thesis is one notable exception.

typical cases where both X and Y are present, as studying cases where we a priori know the mechanism cannot be present tell us little about how the mechanism functions when it is present.

The argument in this paper proceeds in two parts. Given that the existing literature does not clearly distinguish between different variants of PT, we first develop the commonalities and differences across the three variants. The three variants are a case-centric type of PT that we term explaining outcome PT, and two theory-centric variants: theory-testing and Theory-building PT. Regarding commonalities, we focus our attention on the elements of PT that differ from other research methods, be they cross-case or within-case (e.g. congruence). This includes how we conceptualize concepts and mechanisms (set-theoretically), along with how we make inferences (Bayesian logic). The three variants differ on a number of dimensions, including: whether they aim to build or test theories about mechanisms; whether they are 'case-centric' or have generalizing ambitions beyond the single case (theory-centric); and the types of inferences the variants can be used to make.

After delineating the three variants of PT, we then turn to the heart of the article - developing case selection strategies for each of the variants. When using explaining outcome PT, we contend that cases are chosen based on their historical importance. Outcome are understood in a broader, more holistic fashion, such as the Cuban Missile Crisis, or the outbreak of the First World War, instead of as 'cases of' a delimited theoretical phenomenon such as deterrence bargaining or crisis decision-making. One example of this type of study is Allison's classic 'Essence of Decision' (1971).

Within the two theory-centric variants of PT, case selection strategies as a general rule focus on selecting typical cases. Strictly speaking, both theory-centric PT variants only enable us to make *within-case* inferences about causal relationships; in particular how a particular X (or conjunction of X's) contributes to produce an outcome in a single case. Therefore, if we have the ambition to make inferences about causal relationships beyond the single case that we investigated using PT, we have to nest our PT case studies into cross-case research designs – making case selection techniques vital. Cross-case designs include everything from small-n, focused comparisons, classic comparative case studies (e.g. most-different systems design), QCA analysis, to large-n, regression-based research.

A typical case is one that is a member of the set of X, Y and the relevant scope conditions for the mechanism. Scores for X and Y in typical cases must surpass a qualitatively defined threshold that tells us whether a case is a member of the set of the given theoretical phenomenon (e.g.

democracy or peace). We put forward the argument that selecting pathway cases, where scores on other causes are controlled for, is a less relevant strategy when we take the study of mechanisms seriously in PT. The focus of our attention in PT is on how a given mechanism contributes to produce Y, and as PT offers us tools that enable our empirical tests to isolate the workings of particular mechanisms, the scores of cases on other potential confounders become less relevant. We also discuss the role that deviant cases play in theory-building PT. Our argument is that deviant cases are only relevant when we combine a PT study with a comparative analysis that contrasts the deviant case with a typical case (or cases).

2. What is PT and what can it be used for?

Given the lack of clarity regarding what PT actually is, and the lack of differentiation into different variants, we first define what PT is and develop the three variants before we turn to case selection strategies in each of the variants in the next section. The three variants are theory-testing, theory-building, and explaining outcome PT.

Commonalities shared by the three variants

What all three variants share is the core ambition that defines the essence of what PT is; the ambition to trace causal mechanisms using within-case analysis. All three variants also share a range of commonalities, including the use of deterministic theorization, the use of Bayesian logic to make *within-case* inferences about the presence/absence of each of the parts of a causal mechanism (Bennett, 2008b), and a mechanismic understanding of causation, focusing on the process whereby causal forces are transmitted through a series of interlocking parts of a mechanism to produce an outcome (Glennan 1996; Bunge 1997; Hernes 1998). A causal mechanism transfers causal forces from X to Y in a fashion that should leave observable fingerprints that can be studied empirically.

Bayesian logic enables causal inferences to be made without investigating variation (Bennett 2006). For example, in PT empirical tests are developed where predictions are put forward about what evidence we should find in the empirical record if a given part of a causal mechanism is present. Existing ideas about test strength (Van Evera, 1997) are Bayesian-compatible. *Uniqueness* relates to predictions of what evidence we should find that cannot be explained by plausible alternatives (in Bayesian terminology this is the likelihood ratio). Further, what evidence do we have to see for us to say a part of a mechanism is present? This relates to the *certainty* of

predictions. Taken together, Bayesian updating enables us to make inferences about the presence/absence of parts of mechanisms by comparing our predictions with what we actually find in the empirical record without the need to investigate variation or difference-making by causal factors – a point we return to in section 3. The strength of the inferences we can make depends upon the strength of the empirical test, along with the accuracy of the evidence gathered.

All three variants of PT share a theoretical understanding of mechanisms as invariant; either the mechanism is present in a case or not. This implies that set-theoretical understandings of concepts and causal relationships are utilized. We argue that when we conceptualize our causal mechanism and our X and Y, we should *not* treat our theoretical *concepts* as variables; instead they are conditions/outcomes. PT is based on a deterministic and mechanistic logic which implies that we do not study variation but rather whether our theoretical concepts are present in our case. A set-theoretical understanding of theoretical concepts involves focusing on whether a given case is a member of the set of the concept or not. Here the focus is not upon defining the full variation of the concept (differences in degree), but instead is on defining the concept itself (i.e., the concept is present or not present) (differences in kind).[3]

For example, if we are studying a democratic peace mechanism, we are interested in studying democracy and how it contributes to produce peace (outcome). Defined as a condition, only the characteristics that demark the set of the concept of democracy are defined. When conceptualizing democracy, we would need to have a full definition of the positive pole (democracy), but the negative pole (autocracy) would not have to be defined. Basically, the negative pole (autocracy) is analytically irrelevant since studying autocracy does not tell us anything about *how* democracy produces peace. Therefore we would conceptualize democracy by focusing on the positive pole (the characteristics associated with democracy), where the absence of characteristics is simply anything but democracy (~D). Outcomes are also conceptualized in the same manner, with the focus being on the concept and its presence or absence, with and the negative pole merely defined as the absence of the outcome. If on the other hand we are studying the *difference* democracy makes upon the incidence of conflict (Y defined as peace or war), we would need to conceptualize both poles of the regime type (democracy and autocracy).

Ideas about asymmetric causation as they are used in set-theory within comparative

3 The difference between difference in degree and kind can be illustrated with an example of being pregnant. You cannot be a little bit pregnant (difference in degree); either one is pregnant or not (there is a difference in kind). This does not mean that once in the set of pregnancy that you cannot be more or less pregnant. The point is that the matter of degree of pregnancy (for instance between six weeks and eight month) has to be understood as being variation within the set of the concept.

methodology are also applicable in PT. In set-theory, theorizing that developed countries are a subset of democratic countries makes no claims about whether less-developed countries are democratic or not (Ragin, 2008: 15). What is crucial to notice is that the causes of a given outcome are often very different than the causes of its negation. We operate with similar ideas in PT, where the mechanisms that produce an outcome are expected to be very different from those that produce its negation. For instance, the mechanisms leading to war would be expected to be very different than the mechanisms that contribute to peace. Therefore, it makes little sense to compare the mechanisms that produce an outcome with those that produce its negation, given that they are expected to be very different. Therefore, in PT we focus our analytical attention in PT on the mechanism between a given X and the outcome; everything else is analytically irrelevant.

Difference between the three variants

In contrast to the existing literature, where PT is usually treated as a singular method (Gerring, 2007: 172-185; Checkel, 2008:263; Bennett, 2010: 208), we believe that there exist three variants of PT analysis; two of which are theory-centric (theory-testing and theory-building), while one focuses on explaining the outcome in a particular case.

What differentiates the three variants is:

- whether they focus on the case itself, or focus on building/testing more generalizable theories;
- the types of research questions that they are able to answer;
- whether they aim to build or test theorized causal mechanisms;
- their understanding of the generality of causal mechanisms (from systematic mechanisms expected to be present in a population to case-specific mechanisms);
- the types of inferences being made, where theory-testing/building variants are only able to make inferences about the presence/absence of a mechanism in a single case, whereas explaining outcome PT enables inferences about the sufficiency of the explanation in the single case to be made by adopting a different understanding of science;
- The types of case selection strategies used in the different variants, especially between case-centric and theory-centric variants.

The distinction between case- and theory-centric PT has so far been neglected in the

methodological literature, with the result being unclear recommendations on what PT can be used for and which strategies for case selection should be chosen in different research situations. Yet this distinction captures a core ontological and epistemological divide with the social sciences, where we find on the theory-centric side both neo-positivist and critical realist positions (Jackson, 2011). The ambition in theory-centric variants is to build generalizable theories about mechanisms that can travel across cases, within the context in which they are predicted to operate. Causal mechanisms are treated as middle-range theories, and are expected to be present in a population of cases when the causal conditions that trigger them are present, and they are within the proper scope conditions.

In contrast, case-centric scholars who employ what we term explaining outcome PT operate with a very different understanding of the social world, viewing it as very complex and extremely context-specific. In this understanding of the world, generalizations become difficult, if not impossible, meaning that the ambition becomes to account for particularly puzzling and historically important outcomes. The core of this position is well expressed by Evans, who writes, 'Cases are always too complicated to vindicate a single theory, so scholars who work in this tradition are likely to draw on a mélange of theoretical traditions in hopes of gaining greater purchase on the cases they care about.' (1995: 4).

Figure 1 illustrates the different research questions that the three variants aim to tackle, illustrating the difference between building and testing on the theory-centric side, and the broader aim to explain a particular historical outcome in explaining outcome PT.

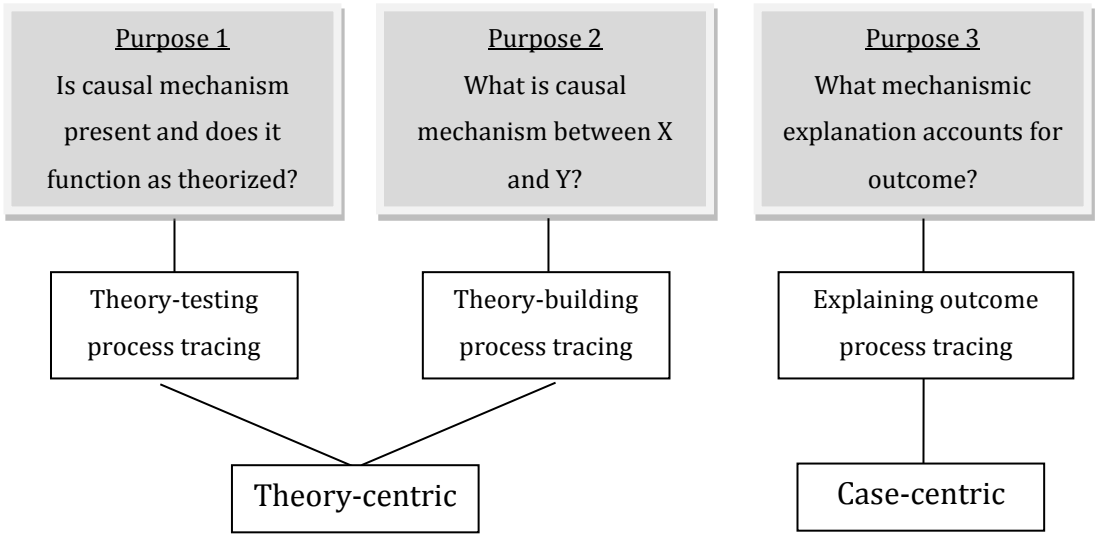


Figure 1 - Three different variants of process tracing and their analytical purposes

We now turn to describing what distinguishes the three variants from each other.

Explaining outcome PT

The goal of many PT studies is to explain a particular historical outcome, working backward from the known outcome to uncover the causal mechanisms that can *sufficiently* explain the outcome. Outcomes here are not understood to be a 'case of' some theoretical concept (e.g. a war), but instead are understood in a much more inclusive, holistic fashion as the Cuban Missile Crisis, or World War I. While the Cuban Missile Crisis could be understood as a case of narrower theoretical phenomena such as deterrence bargaining, crisis decision-making, or bureaucratic implementation, in explaining outcome PT the focus would instead be on explaining the 'big and important' things in the Cuban Missile Crisis, adopting a more holistic view of the particular case (see e.g. Allison and Zellikow 1999).[4]

Underlying explaining outcome PT is a very different philosophical understanding of science than what we know from neo-positivist and critical realist understandings; this includes philosophical positions such as analyticism (monism), and pragmatism (see Jackson, 2011; Humphreys, 2010). Irrespective of the philosophical path to this position, case-centric researchers agree that the social world is very complex, multi-factored and extremely context-specific. This complexity makes the ambition of producing knowledge that can be generalized across cases difficult, if not impossible. Instead, the analytical ambition is to account for particular historical outcomes by using theories in a much more pragmatic fashion (Humphreys, 2010: 269-270). Instead of parsimonious and systematic mechanisms, case-centric researchers contend that it makes little sense to distinguish between systematic and case-specific parts given the impossibility of generalizing in the complex social world. This position is found in the burgeoning literature on topics such as eclectic theorization (Sil and Katzenstein 2010), and pragmatism as a research strategy (Friedrichs and Kratochwill 2009). For instance, eclectic theorization as employed by Sil and Katzenstein only makes sense when the ambition is to account for a particular outcome, as an 'eclectic' combination

4 However, Allison and Zellikow's conclusions do discuss the *lessons* that are potentially applicable to other comparable cases; lessons that can be understood as *potentially* systematic mechanisms that can be investigated in further research in other cases (Ibid). The conclusion however is that the inclusion of non-systematic mechanisms in explaining outcome process tracing studies makes it impossible to nest this type of process tracing case study explicitly in a mixed-method research design (Rohlfing, 2008: 1494-1495).

The aim of explaining outcome PT is to establish an explanation for why an outcome has been produced in a specific case. More precisely, the aim is to provide a 'minimally sufficient' explanation where there are no redundant parts in the explanation (Mackie, 1965). In most circumstances a single theorized causal mechanism is not sufficient to explain the outcome, and therefore it has to be supplemented with new parts from other theories in order to achieve minimal sufficiency, making eclectic theorization and pragmatism relevant.

Explaining outcome PT is an iterative research strategy that aims to trace the complex conglomerate of systematic and case-specific mechanisms that produced the outcome in question. Sufficiency is confirmed when it can be substantiated that there are no important aspects of the outcome that are unaccounted for by the explanation (Day and Kincaid, 1991).

Theory-testing PT

In theory-testing PT we deduce a theory from the existing literature and then test whether there is evidence that a hypothesized causal mechanism is actually present in a given case (George and Bennett 2005; Bennett 2008a, 2008b; Checkel, 2008). The analytical ambition is to test whether the observable implications of the existence of a more general, parsimonious theorized causal mechanism are present in a particular case, comparing the predictions of what evidence we should find with what we actually find in the empirical record.

This variant of PT is often used when a robust empirical correlation between an X and a Y has been found in previous research and we can deduce a mechanism from existing theorization, but we are unsure whether there is an actual causal mechanism linking X and Y.

The belief that theory-testing PT can be used to test two competing theories against each other is widespread but erroneous in most research situations (see George and Bennett 2005: 214-215; Zaks 2011).[5] In the complex social world, most outcomes are the product of multiple mechanisms acting at the same time. The inferences that can be made with theory-testing PT are however restricted to claiming that a mechanism was present in the case and that it functioned as expected. No claims can be made about whether the mechanism was the *only* factor that resulted in outcome Y occurring – in other words, we cannot claim sufficiency based on a single theory-test PT unless

5 - The only exception is when it is possible to conceptualize and operationalize two competing mechanisms in a manner where they are composed of the same number of parts, each of which is the polar opposites of each other and is mutually exclusive. Yet this is a rare situation in social research, and more common is the situation where theories are acting at the same time. Therefore we can only make inferences that a mechanism was present, cognizant that there can be other mechanisms at play in case that also contribute to producing Y.

we jump over the philosophical divide and utilize explaining outcome PT.

Theory-building PT

Theory-building PT has the ambition is to build a theoretical explanation from the empirical evidence of a particular case, inferring that a more general causal mechanism exists that should also be found in the population of the phenomenon. Theory-building PT therefore has ambitions beyond the confines of the single case. In its purest form, theory-building PT starts with empirical material and uses a structured analysis of it to detect the observable manifestations of a plausible causal mechanism whereby X is linked with Y. Theory-building PT can be used both to find causal mechanisms between X and Y, and in situations where we are in the dark regarding what the causes of Y are.[6]

3. Case selection techniques in the three variants of PT

Having explored the commonalities and differences between the three variants on a range of dimensions, we now focus our attention on how they differ regarding case selection strategies. In explaining outcome PT, case selection is based solely on the historical importance or the puzzling nature of a specific outcome, with the outcome understood in a more holistic fashion as the big and important things that occurred in a particular event. In contrast, in theory-testing PT we choose only typical cases based on the argument that why should we choose to test whether a mechanism is present in a case where we a priori know it is not present. Finally, case selection strategies in theory-building PT are also focused primarily on choosing typical cases, although the two types of deviant cases are relevant when combined with cross-case comparison with typical cases.

Explaining outcome PT – the case, the whole case, and nothing but the case

Case selection strategies in explaining outcome process tracing are driven by a strong interest in accounting for a particular interesting and/or historically important outcome. The outcome is not viewed as a 'case of' something, but instead is a particular event that is expressed as a proper noun.

6 - We are more skeptical regarding the utility of theory-building PT for the second goal. See p. ??? more.

Examples of this type of study in the literature include Allison's classic study of the Cuban Missile Crisis (Allison and Zellikow 1999), Layne's study of US grand strategy towards Western Europe after WWII (Layne 1996), and Schiff's analysis of the creation of the International Criminal Court (Schiff 2008). For instance, Allison justifies the choice of the Cuban Missile Crisis by writing, 'The Cuban missile crisis stands as a seminal event. History offers no parallel to those thirteen days of October 1962, when the United States and the Soviet Union paused at the nuclear precipice.' (Allison and Zellikow 1999: 1).

While case selection in explaining outcome PT can resemble the selection of extreme cases, it is vital to underline that a case like the Holocaust, when understood in a more holistic fashion, is not just a 'case of' a theoretical concept like genocide. In explaining outcome PT the ambition would be to craft a sufficient explanation that captures the unique character of a specific (horrific) historical event like *the* Holocaust. We choose the case because it is the Holocaust, or the Cuban Missile Crisis; cases that in and of themselves are historically important to understand the causes of.

The findings of explaining outcome PT cannot be generalized to other cases for two reasons. First, the case itself is unique given our broader conceptualization of outcomes (*the* Cuban Missile Crisis instead of a narrower theoretical phenomenon like a case of deterrence bargaining). Second, given the inclusion of non-systematic parts and case-specific combinations of mechanisms in our explanations, the actual explanation is also case-specific. In Allison's study, he focuses on explaining a set of central questions about the case using three different theoretical lenses, resulting in a composite, case-specific explanation.

Therefore, the conclusion is that when we in Evans' words care deeply about a particular case, the case selection strategy we employ is one based solely on the puzzling nature and/or historical importance of a case. Cases are therefore not 'cases of' a generalizable theoretical phenomenon.

We now turn to developing guidelines for case selection when using the two theory-centric variants of PT. Before we discuss the specifics, it must first be pointed out that, logically, PT case studies can be undertaken before or after cross-case analysis. However, in our opinion, both theory-centric variants of PT should only be used after cross-case analysis has been undertaken using comparative methods. In our opinion we should never engage in PT before cross-case analysis has been done using other methods.

Before cross-case PT?

The existing literature suggests that PT before cross-case analysis can be used to define the relevant population of a given phenomenon (Ragin, 2000), or to find its potential causes (Goertz, 2008; Rohlfing, 2008). However, if we take the study of causal mechanisms seriously, we argue that we should never engage in before cross-case PT. In our opinion, this is in effect like starting to build an expensive building without any knowledge of what it will be used for. Our argument here is that PT is not the best tool in the methodological toolbox for either before cross-case purpose.

Regarding the first before cross-case use, how can we learn anything about the population by engaging in an in-depth study of a single case? Recommending this type of exploratory case study is based on a misunderstanding of PT, where the method is seen as a form of descriptive, narrative case study instead of a method focused on investigating causal mechanisms between X and Y. For example, if we want to test a causal mechanisms dealing with power in policy-making processes within the EU, we would not blindly select a policy area. Instead, we would first *compare* policy-areas across a range of potentially relevant theoretical criteria (such as salience, number of veto players, size of distributive implications, etc) to identify similarities and key differences that can be expected to affect whether the theorized mechanism is present and how it operates.

By selecting blindly, we might be lucky that we chose a (relatively) 'typical' policy-area like environmental policy, or we might have chosen a highly unique area like the Common Agricultural Policy. The key point here is that without *comparing* across policy-areas (at the very least in a descriptive fashion), we know nothing about the population of the phenomenon, as we have no idea about what is idiosyncratic and what is typical across cases. Therefore, if we find the causal mechanism in the case, we have no idea about whether what we have found is unique to the case (case-specific/non-systematic) or whether we should expect to find the mechanism more widely in the population.

Second, methods such as small-n comparative studies offer more bang-for-the-buck in relation to detecting potential causal conditions than PT does. Using a most-different systems design (MDSD), it is relatively easy to compare two or more cases that differ on the outcome to detect which causal condition(s) vary, and which are similar in both cases and thereby are eliminated as potential causes. In contrast, a theory-building PT case study would involve looking for the observable manifestations of an unknown causal mechanism that was triggered by an unknown cause – in effect blindly groping for mechanisms without any clue about what we should be looking for in the empirical record. While this can be done, the magnitude of the research effort

involved is reflected in the fact that there are very few good examples (if any) in the literature of PT succeeding in this type of research situation, whereas there are numerous examples of theory-building small-n comparative analysis that detect plausible causal conditions.

We therefore recommend first that scholars *learn something about the population* of a given phenomenon before they engage in selecting cases for PT. This can even be in the form of descriptive comparative research, where the scholar gains enough knowledge about the cases to be able to make informed choices about whether cases are typical or not (King, Keohane and Verba, 1994: 34-74; Gerring, 2012). In the above example of EU policy-making, this could involve reading descriptive accounts of the EU policy-making process, and how insiders or observers perceive what is similar across policy-areas, and what key differences are.

Further, as we will develop further below, we recommend using cross-case methods such as small-n comparisons to detect potential causal conditions. After candidate X's are detected using these tools, we can then engage in theory-building PT to uncover the causal mechanisms that link X and Y.

The conclusion is that we should never engage in pure before cross-case PT. At the very least we should have some knowledge of the population of the phenomenon, meaning that PT case selection should always be thought of as after cross-case PT. Anything else is blindly groping in the dark.

After cross-case PT research

Case selection strategies vary according to which theory-centric variant of PT is being used, which itself is a function of the type of research question one is faced with (see figure 1, section 2 above). When we want to know how X contributed to producing Y, and when it is possible to formulate a plausible causal mechanism between them, we choose theory-testing PT. If we know X and Y but are in the dark regarding how X contributes to causing Y, we choose theory-building PT. If we know X and Y, but have a deviant case where X did not contribute to produce Y, we choose theory-building *combined* with comparative methods to detect where the causal mechanism from X broke down, and whether either key causal conditions were omitted and/or whether the scope conditions for the operation of the mechanism were not present. Finally, if we are in the dark about both the causes of Y and the mechanism whereby the unknown causes(s) were linked with the outcome, one can also use theory-building PT – although this is not a strategy we recommend given that there are

more effective tools at detecting new X's in the methodological toolbox than PT.

We review in the following case selection strategies for both the theory-testing and theory-building variants of PT. Common to both is the importance of qualitative thresholds of concepts (i.e. membership or non-membership in the set of relevant causal conditions and the outcome) (see section 2). We argue that typical cases, defined as '…representative of a population of cases (as defined by the primary inference)' (Gerring, 2007: 96), are most commonly used in both theory-centric variants of PT, whereas deviant cases (where either X or Y is not present) also serve certain more limited theory-building purposes.

Theory-testing PT

We put forward the argument that typical cases are the only type of cases that should be selected for testing whether theorized causal mechanisms are present in a case. This argument contrasts existing recommendations in three ways. First, while our general recommendation that we *choose typical cases* is found in some parts of the literature (Lieberman 2005), what distinguishes our approach is that we contend that typical cases are not those on or near the regression line (low residuals), but only those that are members of the sets of both the causal condition (or conjunction of conditions) and outcome. Second, *what a relevant typical case looks like is not affected by the type of causal claims being made*. Simply put, if we take the study of mechanisms seriously, choosing a typical case should be solely based upon values of X, Y and scope conditions, and not other potential causal conditions. Third, in contrast to existing recommendations (Lieberman 2005; Rohlfing 2008), we suggest that we should engage in multiple theory-test case studies instead of being satisfied with a positive finding in one case. Due to potential multiple causal pathways between X and Y (Gerring, 2010), we cannot infer based on finding a causal mechanism in one PT theory-testing of a typical case that the same mechanism is also present in all other typical cases within the population.

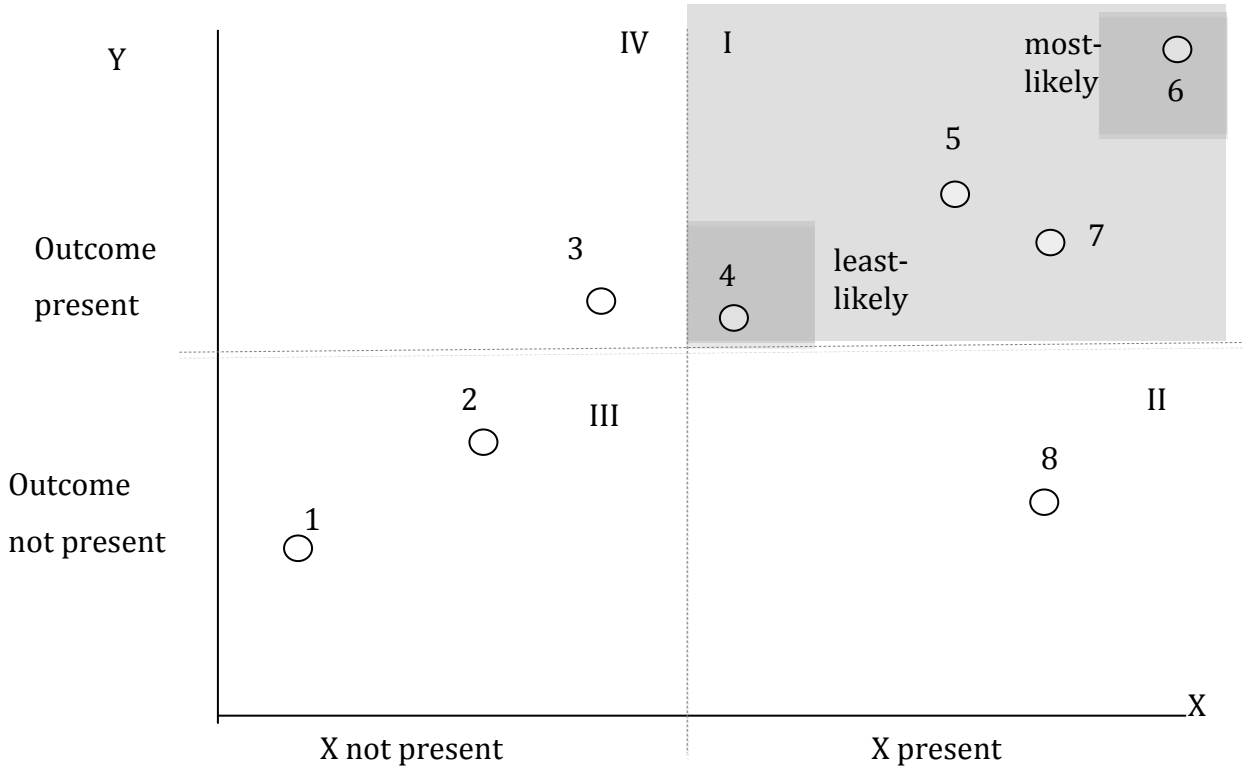
Figure 2 illustrates the core of our argument relating to theory-testing PT. The figure illustrates a simple X:Y causal relationship.[7] Qualitative thresholds demarcating set membership of the causal condition (X) and the outcome (Y) are depicted with dotted gray lines. These thresholds

[7] - The figure can also be adapted to capture conjunctions of causes, meaning that instead of just X, set membership for a conjunction (e.g. X1*X2*X3) would include cases that are members of the sets of both X1, X2 and X3. A case in zone 1 would therefore be both a member of the set of each of these conditions in a given conjunction and the outcome. If one of the conditions in the conjunction was not present but the outcome was present, the case would be found in zone IV.

demarcate four zones of cases, where for theory-testing PT only zone I is relevant, where both X and the outcome are present. Note that if information about differences of degree within sets is available, we can also distinguish scores for cases within the zones (see below for more).

Not captured in the figure, but also important, are whether the relevant scope conditions (or context) that allow a given mechanism to function are present or not (Falletti and Lynch, 2009). Scope conditions are defined as the '…relevant aspects of a setting (analytical, temporal, spatial, or institutional) in which a set of initial conditions leads…to an outcome of a defined scope and meaning via a specified causal mechanism or set of causal mechanisms.' (Falletti and Lynch, 2009: 1152). If we have a car causal mechanism (that produces forward movement), we might theorize that the scope conditions in which it can be expected to operate include the presence of oxygen and relatively level ground. If we throw the car in a lake, while the car might be in perfect shape it will still not work, as it is outside of the scope conditions in which it will run. A causal mechanism linking economic development that operates in one regional context (East Asia) might not necessarily function in another context (Latin America). When choosing typical cases, we therefore also need to take into consideration whether the relevant scope conditions are present.

Figure 2 – Case selection in theory-testing PT



Returning to figure 2, the core of our argument is that only typical cases within zone 1 should be chosen in theory-testing PT. Cases in zones III and IV are irrelevant for the simple reason that given that X is not present, we know a priori that the causal mechanism linking X to Y will not be present. Cases in zone II are also not relevant for theory-testing PT, given that we also know that the causal mechanism is not present. We argue below that cases in zone II are deviant cases that are extremely interesting for building theories about causal relationships (conditions and mechanisms), given that they can for example tell us something about the scope conditions under which a given causal mechanism functions (and importantly here, does not function). But again, why test whether a theorized causal mechanism was present when we know a priori that it was not present?

Therefore, when engaging in *theory-testing PT*, we should *only choose cases in zone I*. If we only have information about cases that enable us to assess set membership or not (crisp set membership), all cases within zone I are typical cases. However, sometimes we have information about both differences in kind (qualitative thresholds) and differences of degree both as regards causal conditions (X) and the scope conditions that are not depicted in the figure. These differences of degree are what are captured in fuzzy-set membership scores (see Ragin, 2000, 2008).

We can utilize this additional information of differences in degree within zone I to categorize cases according to whether they are cases where we expect the mechanism is most- and least-likely to manifest itself. Most-likely cases are those with high scores of both X and Y within zone I, whereas least-likely cases have low values but are still within zone I. As regards least-likely, it is important to note that it should still be plausible that it can exist in the case (at least theoretically), meaning that it is still within zone I (Mahoney and Goertz, 2004). When determining whether a case is most, least-likely or 'just' typical, scores for both the causal and scope conditions are relevant.

Information that enables us to identify most and least-likely cases for the operation of a given causal mechanism provides us with stronger analytical tools for selecting cases that match the specific research situation we are in. If we initially are very skeptical regarding whether a given causal mechanism even exists (low prior confidence), we suggest that one should choose a most-likely case for a theory-test. Here a positive result will significantly update our confidence in the mechanism in a positive direction (confirmation). However, finding a given mechanism in case 6 does not necessarily mean that the same mechanism is present in cases 5 and 7.

In contrast, if we have a relatively high degree of prior confidence in the mechanism, we might select a least-likely case. If we find the mechanism here, we can make strong claims about

causality based on the Bayesian-inspired 'Sinatra inference' (Levy, 2002: 144). However, not finding the mechanism in a case where we a priori were not very confident about finding it does not enable us to update our confidence in whether the mechanism was actually present or not, making this a high-risk strategy. If we do not find the mechanism in a least-likely case, we do not really know whether the mechanism is actually present in the rest of the population, whether it is just very difficult to detect in least-likely cases (meaning it is there but that our empirical tests were unable to capture it, i.e. a false negative), or whether it just not present in cases where we a priori least expect it to be present (but at least possible to exist).

We now turn to a discussion of the three ways in which our recommendations differ from existing ones.

Are typical cases only in zone I?

The recommendation to restrict theory-tests to typical cases within zone I conflicts with the still widely-followed prescriptions for case selection put forward by Lieberman (2005). According to Lieberman, a typical case is one with small residuals in relation to a regression-based analysis of X:Y correlations (Lieberman, 2005: 444). He writes that '…cases that are on, or close to, the 45-degree line (plotting actual dependent variable scores against regression-predicted scores) should be identified as possible candidates for in-depth analysis.' (Ibid). This implies that in figure 2, all cases close to where a regression-line would appear are relevant – for example a line running from case from 1 to 6. Lieberman goes even further to contend that our confidence in a causal relationship would be increased even further if we select two or more on-liers that have a wide range of observed scores on the outcome (Ibid), meaning that he would prescribe that we select cases 1 and 6 for theory-testing PT.

However, cases 1, 2 and 3 all are instances where we know a priori that a causal mechanism will not be found between X and Y! Lieberman's recommendations for case selection therefore ignore the very purpose of theory-*testing* PT, which is to investigate whether the hypothesized causal mechanism was actually present in a case where it at least theoretically can be present. It is here that the qualitative thresholds discussed earlier in the paper play a critical role in determining whether we can expect a mechanism to be present or not. For the mechanism to be present, case scores for both X and Y have to surpass the qualitative threshold delimiting members of the set of the condition and outcome. Lieberman's guidelines do however make some sense if we restrict

selection to cases within zone I; a point we return to below regarding multiple theory tests.

Typicality does not depend upon case membership in other causal conditions

What a typical case looks like does not depend upon what types of claims about causal relationships we are making.[8] In the following we show that in contrast to existing recommendations, we should classify cases as typical that can be used for theory-testing *solely* based on values of X, Y and scope conditions, ignoring membership scores of cases for other causal conditions. We focus our attention on the strategy of using pathway cases suggested by Gerring and Seawright 2007 and Schneider and Rohlfing forthcoming.

Gerring and Seawright use the term pathway case to describe a case that in their opinion best allows us to focus on causal mechanisms between a given X and Y.[9] A pathway case is defined as '…one where the causal factor of interest, X1, correctly predicts Y's positive value (Y=1) while all other possible causes of Y (represented by the vector, X2) make "wrong" predictions…If X1 is…a sufficient cause of Y, then it is these sorts of cases that should be most useful for tracing causal mechanisms.' (2007: 125). While Schneider and Rohlfing use the term unique membership, their logic is the same.

We contend that scores for other possible causes are less relevant for theory-testing PT case selection, given that we are *not* interested in investigating the 'causal effect of one factor' (Gerring 2007:122), but instead are focusing on whether a causal mechanism linking X1 to Y has left the predicted observable fingerprints in a case. If we want to investigate the difference in the value of Y that the value of X1 makes, we would want to have a design that approaches a medical experiment, where the effects of X1 are isolated from the effects of potential confounders. Yet when we engage in theory-testing PT, our analytical attention is in contrast focused on the mechanism itself and how it transfers causal forces from X to contribute to produce Y.

There are two complementary reasons for why membership scores in confounding causes are

[8] - Mechanisms should be theorized as deterministic causal relationships given that we investigate them in PT using invariant single case studies. Mahoney writes that, 'At the individual case level, the ex post (objective) probability of a specific outcome occurring is either 1 or 0; either the outcome will occur or it will not…single-case probabilities are meaningless.' (2008: 415-416). However, probabilistic relationships can be re-conceptualized in a more deterministic fashion to enable PT. For more on this point, see Beach and Pedersen, in press.

[9] - Although they focus on the use of pathway cases for building theories of mechanisms, the same strategy should logically apply for theory-testing PT.

less relevant for PT case selection. First, if we conceptualize and operationalize our tests of the X1 causal mechanism properly, choosing a case where X1 and X2 are both present should not matter, in that if there actually is a mechanism operating between X1 and Y, we would find the predicted observable manifestations of its existence in the case *irrespective* of whether there were other mechanisms also in action (see p. ??? above). Empirical tests of each part of a mechanism should be formulated in a manner that eliminates potential alternative explanations of finding the evidence (i.e. uniqueness). When unique predictions are made, the presence of other causal conditions in a given case is analytically irrelevant. What difference does it make when we are tracing a mechanism leading from a causal condition to the outcome that another causal condition also was present when our empirical tests enable us to isolate the workings of the mechanism itself from other causes? For instance, if we are testing a mechanism linking economic development with democracy in a situation where we also believe that external pressures could have contributed to democratization, a proper theory-test could develop relatively unique predictions about what evidence we should find in the case that could only be accounted for plausibly if there was an actual development-democracy causal mechanism present and functioning in the case.

Second, case selection strategies that are designed to assess whether a condition and/or mechanism was necessary and/or sufficient are not relevant for PT, as theory-testing PT offers us no tools to determine whether a condition (and/or mechanism) was necessary or sufficient to produce an outcome, in contrast to claims made by Schneider and Rohlfing (forthcoming). PT can only be used to make claims about whether a mechanism is present or not in a given case (see Beach and Pedersen, in press). Again, this means that scores of cases on other factors is less relevant.

For example, claims of necessity do not rule out other causal conditions also contributing to producing the outcome; indeed we should expect multiple causal conditions contribute to produce the outcome in any given case. Yet tracing the causal mechanism linking a necessary condition to an outcome focuses exclusively on whether the predicted observable manifestations of each of the parts of the hypothesized causal mechanism are present or not. A single PT case study cannot prove that a causal condition is necessary in the case unless we heavily utilize counterfactual reasoning, as we cannot control for the non-occurrence of X in a single case. [10] Testing necessity can only be

[10] - In our work on PT, we are skeptical regarding the analytical utility of counterfactuals in studying causal mechanisms. We believe that a combination of Bayesian logic and strong empirical tests eliminates the need for artificial 'variance' that could be created by relying on speculative counterfactual reasoning. For an example of an understanding of PT that relies heavily on counterfactuals, see Collier, 2011.

done when we have variation in X (which we do not have in an invariant PT single case study); instead cross-case comparative methods should be used to determine necessity. Therefore, we argue that typical cases for theory-testing PT would include *all cases* where X and Y are present (along with the scope conditions).

If claims about *sufficient causal relationships* between X and Y are being made,[11] we also contend for the same reasons that one should only concentrate on values of X and Y – thereby disregarding case membership in other causal conditions – in contrast to the arguments put forward by Gerring and Seawright (2007) and Schneider and Rohlfing (forthcoming).

Gerring and Seawright further discuss pathway cases in situations where variables are continuous, suggesting that pathway cases will satisfy 'two criteria: (1) it is not an outlier (or at least not an extreme outlier) in the general model, and (2) its score on the outcome (Y) is strongly influenced by the theoretical variable of interest (X1), taking all other factors into account (X2). In this sort of case it should be easiest to identify the causal mechanisms that lie between X1 and Y.' (2007: 126). Pathway cases are those that are most affected by the value of X1, and are found by investigating the difference in the size of residuals between the reduced-form model (where X1 is excluded) and the full model (where X1 and X2 are present) (Ibid, p. 127).

While this is sound advice if we want to analyze difference-making, if we want to study a causal mechanism between X and Y, case scores on confounders become less relevant. We illustrate this using the same example of the relationship between oil wealth and regime type provided by Gerring and Seawright (2007: 127-130), which draws on an article by Ross (2001). We use the data set employed by Ross in the following. Gerring and Seawright compare the size of residuals between a full-model (containing X1) and the reduced model (with only a series of control variables), they find two candidate pathway cases. This is illustrated in table 1, below.

[11] - Only rarely are claims about sufficiency of a single causal condition made; democratic peace being one notable exception. However, in QCA claims about conjunctions of conditions being sufficient to produce an outcome are commonly made. The arguments we put forward in this section apply both to single conditions and to conjunctions.

Country	Residual reduced	Residual full	Δ Residual
Belgium	.518	.310	.208
Singapore	-1.593	-1.864	.271
Norway	.315	1.285	-.971
UAE	-1.256	-.081	-1.175
Kuwait	-1.007	.925	-1.932

Source: Gerring and Seawright (2007: 129).

Table 1 – An abbreviated table of possible pathway cases between oil wealth and regime type

They suggest that one chooses the cases where the impact of oil wealth can be best isolated from potential confounders (other plausible causes), resulting in the selection of UAE and Kuwait out of a set of cases that include among others Singapore, and Belgium. Norway is excluded as a potential pathway case, as the size of the residual increases when oil wealth is added. However, we suggest an alternative and much simpler route that is at the same time more compatible with the idea that we are tracing mechanisms instead of investigating the difference that X1 makes, thereby making scores of confounding factors less relevant.

Given that causation is assumed to be asymmetric in PT, we need to first decide *which* causal mechanism we are actually interested in studying: one between oil wealth and autocracy (oil curse mechanism), or one between oil wealth and democracy (an oil blessing mechanism?). This is never made clear in Gerring and Seawright's chapter. They tell us that the factor of theoretical interest is oil wealth, but is the focus on the oil curse or oil blessing mechanism?

If we want to study the oil curse mechanism, we would then need to establish qualitative thresholds for the presence of oil wealth and autocracy, followed by differences of degree if we possess this type of information. For sake of illustration in this paper, we depict crisp-set membership scores for each of these cases based upon a rough calibration using Ross' data-set from his 2001 *World Politics* article.[12] This is shown in figure 3.

[12] - In the Ross data-set, the scores for oil in 1997 as percent of GDP are Belgium (2.03), Kuwait (39.4), Norway (17), Singapore (0.1) and UAE (45 – note observation from 1979). In the data a threshold can be drawn somewhere between the values of 5 and 10%. For regime they are Belgium and Norway (10), Singapore (-2), and UAE and Kuwait (-7). A rough threshold can be set at 0. Given the distribution of the cases here, our arguments are robust across different calibrations of the threshold.

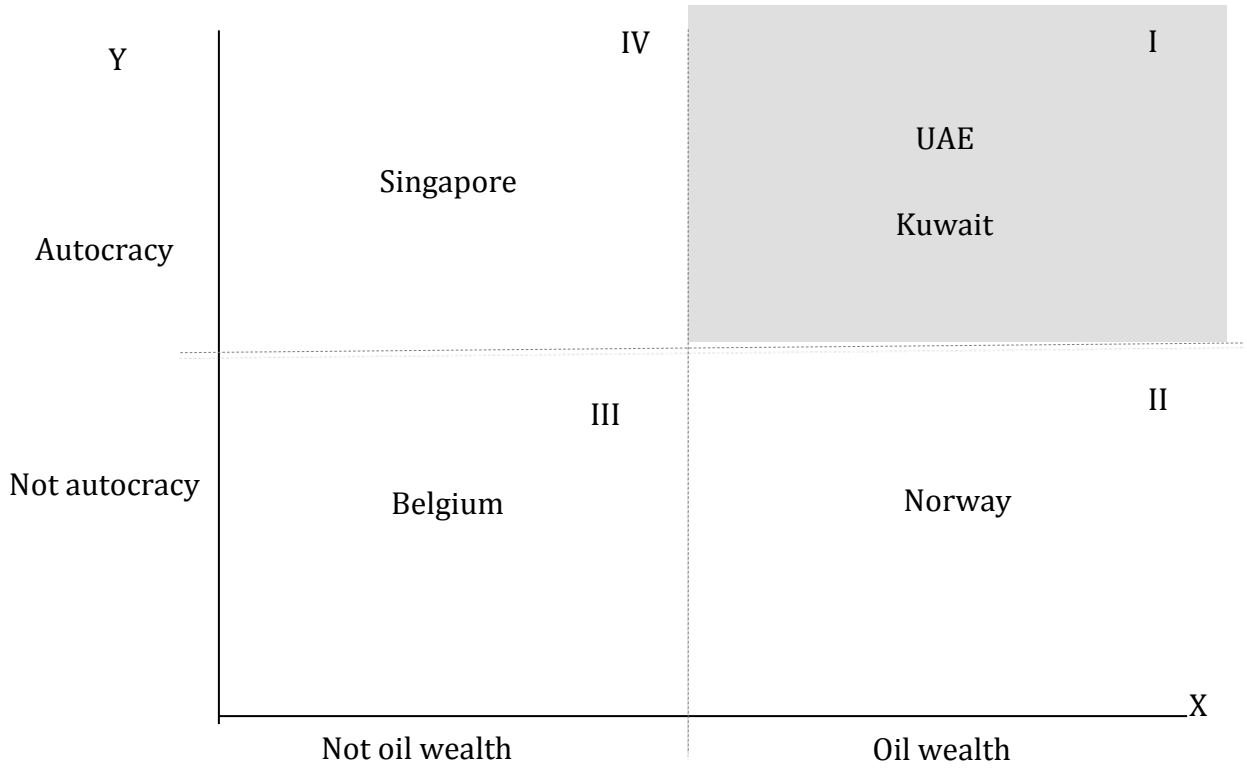


Figure 3 – An example of why confounding causes are irrelevant – the oil curse mechanism.

Using this information, our theory-testing PT would therefore focus only on cases where both X and Y were present, meaning that Belgium, Norway, and Singapore would *not* be typical cases within zone I. As can be seen in figure 3, Belgium is in zone IV, Norway in zone II and Singapore in zone III; we would not expect the oil curse mechanism to be present in any of these cases. The only analytically relevant cases for tracing whether an oil curse mechanism actually exists are cases within zone I (UAE and Kuwait). By adopting qualitative thresholds we thereby reduce the need to control for 'other causes' using pathway cases.

Don't stop till you get enough

Finally, and in contrast to the frameworks proposed by both Lieberman (2005) and Rohlfing (2008), we contend that if we find evidence that increases our confidence in the existence of the theorized causal mechanism in a typical case, we should not be content to stop with one positive finding in a PT case study. Finding a causal mechanism between X and Y in case 5 (figure 2) does not

necessarily mean that the same causal mechanism links X and Y in case 7. Due to potential causal heterogeneity at the level of mechanisms, there might be multiple causal paths between X and Y (Gerring, 2010: 1508-1510). If we find that economic development contributed to produce democracy in the South Korean case through a middle-class mechanism, this does not necessarily mean that economic development contributed to producing democracy in Taiwan through the same causal mechanism.[13]

This implies that we should never be content with a one-shot PT theory-test. To enable stronger claims about the presence of mechanisms across the population (i.e. within zone I), we recommend engaging in a second theory-test. This can involve choosing another typical case, or selecting a least-likely case. In the former, lower-risk strategy, if the same mechanism is found in a second theory-test, we can make cautious claims about mechanisms in the population. As we argued above, choosing least-likely cases can be a high-risk strategy.

If we do not find a mechanism in the first theory-test, we suggest that we should *compare* the chosen case with other typical cases, investigating whether there is something unique about the case. If it is found that the case was idiosyncratic, perhaps due to highly non-systematic (case-specific) factors, we suggest that one chooses another typical case for theory-testing. If however it is found through comparing the chosen case with other typical cases that there is nothing unique about the chosen case, we suggest engaging in theory-building PT on typical cases to attempt to detect whether there is a different causal mechanism than the one theorized linking X and Y.

Theory-building PT

Guidelines for case selection for theory-building PT are slightly more complex than for theory-testing. Case selection strategies for theory-building PT are depicted is figure 4. Whereas theory-testing is a single purpose variant of PT, enabling us to answer the question of whether there is strong evidence suggesting that a hypothesized causal mechanism was present or not in a given case, theory-building PT can be used for three distinct purposes. First and most common is using PT to detect what the causal mechanism is between X and Y (zone I). Here typical cases are chosen. A second use of theory-building PT is that we can use it in *combination* with comparative methods can be used to analyze deviant cases within zone II, where the focus is on detecting either omitted

[13] - Note that all of our examples are purely for heuristic purposes in this version of the paper. We will naturally link the examples more with existing substantive research in the final article.

causal conditions that together with X are necessary to be present in order to trigger the causal mechanism,[14] or scope conditions that can be detected by comparing cases in zone I and II. The PT component of this type of research involves tracing the mechanism from X until it breaks down, which helps identify candidates for condition(s) that can have led to break down that can then be assessed by comparing cases in zone I with zone II. Finally, while we recommend choosing other more cost-effective methods for the purpose, theory-building PT can be used to detect new causal conditions and mechanisms linking them with Y in cases within zone IV.

Typical cases in zone I - building a theory of the causal mechanism between X and Y

When we know that there is some form of causal relationship between X and Y, theory-building PT can be used to detect causal mechanisms between them. Here we want to select typical cases that are within zone I, where the causal condition, outcome and scope conditions are present, meaning that a causal mechanism between X and Y is theoretically possible. As with theory-testing PT, one does not need to select typical cases that resemble pathway cases due to the same reasoning involved in theory-testing (see above).

[14] - Again, assessing necessity is only possible with comparative methods.

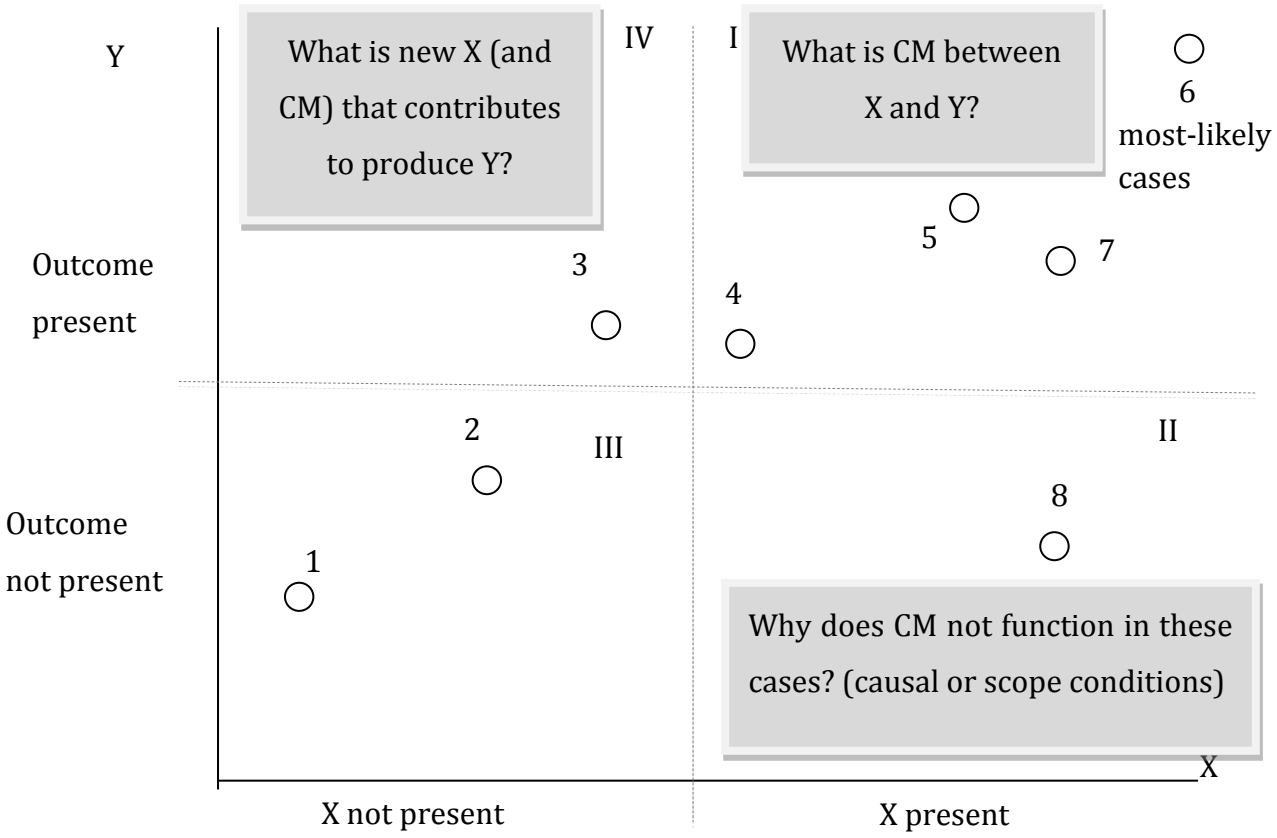


Figure 4 – Case selection in theory-building PT.

Detecting a causal mechanism is easiest when we select a case that based upon our knowledge of the phenomenon looks like a most-likely case (cases within the shaded zone (case 6) in figure 4).[15] An example of this case selection strategy in theory-building PT is found in Janis' classic book on *Groupthink* (1983). The first case that he uses to build the theorized mechanism of Groupthink is the Bay of Pigs fiasco in 1961, where the 'best and the brightest' policy-making group in the early Kennedy administration decided to support an intervention that was doomed to fail from the start due to the faulty assumptions underlying the decision.[16] This can be understood as a 'most-likely' case of flawed small group dynamics, in that the values of X (very cohesive small group) and Y (faulty decision-making process) were very high.[17] The logic behind Janis' selection of the Bay of Pigs case is that it is a perfect storm of factors producing faulty decision-making, and if a

[15] - We contend that one would never select a least-likely case within zone I to build a causal mechanism, given the large risk of simply not detecting the mechanism even though it is (weakly) present.
[16] - For more on how Janis engages in theory-building PT in his book, see Beach and Pedersen, 2012.
[17] - Note that he is not as explicit about whether the scope conditions are present or not, as he focuses on X (which he terms antecedent conditions) in his model (see Janis, 1983: 244).

Groupthink mechanism exists at all, we should expect to detect it in such a case.

However, selecting most-likely cases raises a dilemma, in that should we expect the same mechanism to link X and Y in more representative typical cases such as cases 5 and 7 as in the most-likely case 6? There is not any quick-and-easy fix for this, but we suggest the strategy of iteration, where a mechanism found in case 6 is then tested on more representative typical cases such as 5 and 7 to see whether it is more widely present in the population (see figure 5 below).

What types of inferences can be made after a theory-building PT on a case from zone I? If a causal mechanism is found, it then needs to be tested to see whether it is present in other typical cases. We do not suggest that one immediately goes to testing the theory at the cross-case level, given that the 'theory' we have built is about mechanisms between X and Y, not a theory of a causal relationship between X and Y. It is therefore more relevant to engage in a second round of PT in another case, testing whether the found mechanism is also present in other typical cases. If we then find that the mechanism is also present in another case, we can make the inference that it is probably also present in the rest of the relevant population, enabling us to make claims about how X caused Y through a given causal mechanism.

A key challenge when testing a mechanism in another case is how do we know whether the parts of the mechanism found using theory-building PT are systematic, or are unique to the individual case (Rohlfing, 2008)? If the mechanism we built based on the one case includes non-systematic (case-specific) parts, we would per definition not find it when we test the theorized mechanism in other cases. We can reduce the magnitude of this problem by asking ourselves during our theory-building the question 'can this part of the mechanism at least hypothetically also be present in other cases'. For example, if we build a theorized mechanism that contributes to producing democracy using India as a typical case, and a part of our causal mechanism that we built includes the role played by the Congress Party, based upon our knowledge of the (very limited) population of relevant cases, we would conclude that given the unique role that the Congress Party has played in India, it is a non-systematic part that could not be found in other cases. We should therefore revise the part of the mechanism, re-conceptualizing it as something that could potentially be present in other cases (a systematic part), although in this case it might be impossible to re-conceptualize the Congress Party as a systematic factor given the unique role it has played in Indian politics.

If we are not able to detect a causal mechanism between X and Y in a single case after numerous repeated attempts, there can be two reasons for this: 1) either the case is idiosyncratic, or

2) there is no causal relationship. To determine which of the two is correct requires *comparing* the chosen case with what we know about other typical cases, assessing the values of the case and other typical cases both on X, Y, scope conditions, and a range of other potential causal conditions. In particular, this comparison can shed light on the scope conditions for the functioning of the mechanism, enabling us to assess whether the bounds of the population have been set properly. For sake of example let us say that we do not find the expected 'middle-class growth' causal mechanism linking economic development and democracy in a PT study of what we think is a typical case of third wave democratization. We would then compare this case with what we know more generally about cases within the population. In doing so, we might find that the scope conditions for the operation of the middle-class growth mechanism were not present in the case, suggesting that we should redefine the population to exclude the case. We could for instance have chosen what looked like a typical case (Poland), but found that the scope conditions for the operation of the middle-class mechanism were not present, as Poland was also engaging in parallel state-building processes in comparison to other cases like South Korea or Taiwan. This finding would suggest that the scope conditions for the middle-class growth mechanism are limited to countries not engaging in parallel state-building processes, meaning that we would reduce the bounds of the functioning of the mechanism, thereby reducing the population of cases in zone I.

However, if we cannot detect any significant differences between the chosen case and other typical cases in the population, we can make the cautious inference that there actually is not a causal relationship between X and Y. We would conclude that while there is a correlation between X and Y, there is no underlying causal mechanism linking the two, meaning that the postulated causal relationship is spurious.

Deviant cases in zone II – why does the causal mechanism between X and Y not work in the case?

Cases within zone II are deviant cases, in that X is present but the expected outcome is not present. Here we should otherwise have expected that X would trigger the operation of a causal mechanism that would contribute to the occurrence of the outcome. The basic question here is why the causal mechanism breaks down. There can be two reasons: 1) there are omitted causal conditions from the model that have to be present for the mechanism to operate; and/or 2) there are omitted scope conditions that have to be present for the mechanism to operate. This type of PT is usually not engaged in before theory-tests of typical cases have updated our confidence in the existence of a

given theorized mechanism in a positive direction (within zone I).

Here theory-building PT is not a stand-alone method, but relies on an analytical two-step, where we use PT to trace where the mechanism breaks down. We then use these insights to inform our comparative analysis with cases in zone I to uncover omitted causal and/or scope conditions. The PT component of this type of deviant case study involves tracing the theorized mechanism from X until it breaks down in the case.[18] Using the middle-class mechanism as an example, we might find in a case that while a middle-class was produced, the part of the mechanism where the middle-class demands increased participation was not actually present, with demands for democratization coming instead from elite actors, as happened in many transitions across, Central, Eastern and Southern Europe. We would then compare the single case with what we know more broadly about other typical cases in the population to uncover whether there is a missing causal or scope condition in the deviant case that can explain why the mechanism did not function as predicted. This comparison should preferably include cases where we know that the mechanism itself was present, which is why we suggest that zone II deviant cases are only relevant after one has engaged in successful theory-tests within zone I. Returning to our smoker-cancer relationship, a heavy smoker that does not get cancer (zone II) could be compared with a typical smoker with cancer from zone I, where the comparison could identify conditions such as genetic predispositions present in cases in zone I in figure 4, but not in those in zone II.

Deviant cases in zone IV – Finding new causes with theory-building PT (or mission impossible?)

Cases within zone IV are also deviant cases, although here X is not present whereas the outcome is present. This type of deviant case is relevant to build new theories about the causes of the outcome. While this is one potential use of PT that is advocated by many scholars (e.g. Lieberman, 2005: 443; Rohlfing, 2008: 1510), we contend that theory-building PT is not the most efficient methodological tool to detect new X's. We strongly recommend that scholars first use cross-case methods such as MSSD or MDSD comparisons to detect potential causes of the outcome, followed by either theory-building or testing PT using typical cases of the newly found X, Y and the relevant

[18] - Strictly speaking, the manner in which PT is employed here follows more closely the guidelines for theory-testing PT, tracing a hypothesized mechanism until its breakdown. However, we include it within theory-building PT for two reasons. First, the PT element here has a much more explorative character, soaking and probing in the case to determine why breakdown occurred, gaining insights into what went wrong that can inform the comparative analysis. Second, when combined with comparative analysis of cases in zone I, the overall focus of this type of design is clearly theory-building.

scope conditions.

If one however chooses to blindly grope in the dark to find both a cause (new X) and mechanism and succeeds in finding both, much stronger inferences about the causal relationship between the new X and Y can be made than if small-n comparative methods are used. While we can make strong inferences eliminating potential causes using a MDSD, we can only make relatively weak positive inferences about a causal relationship between the new X and Y based upon finding the expected covariation between new X and Y in the two or more cases in the MDSD. In contrast, tracing backwards from Y to find both a mechanism and X enables us to make much stronger claims about there actually existing a causal relationship between cause and effect (for more see Beach and Pedersen, in press). These claims about a mechanism between the new X and Y then should be tested using theory-testing PT, selecting a typical case of the new X, Y and relevant scope conditions.

4. Conclusions

This paper has developed guidelines for case selection strategies when using each of the three variants of PT. These are depicted in table 2. We see first that case selection strategies are not relevant when we are engaging in explaining outcome PT due to the broader conceptualization of outcomes that is a product of the different understandings of case study research (and science itself) underlying this variant of PT. Here we simply select historically important cases because they are for instance the First World War, not a 'case of' failed deterrence or crisis decision-making.

	Research purpose	Case selection strategy
Explaining outcome PT	- Accounting for the particular historical outcome (crafting a minimally sufficient explanation).	- choose historically important cases. - cases are not a 'case of' something. Instead they involve broader conceptualizations of 'big and important things' in particular historical events such as the Cuban Missile Crisis.
Theory-testing PT	- testing whether there is a hypothesized causal mechanism between X and Y	- choose typical cases (members of set of X, Y and scope conditions – zone I).
Theory-building PT	1) detect causal mechanism between X and Y 2) detect omitted causal and/or scope conditions 3) build new theory of X and mechanisms that contribute to produce Y	1) choose typical cases (members of set of X, Y and scope conditions – zone I). 2) choose deviant cases where X present but Y not present (zone II). 3) choose deviant cases where Y present but existing X not (zone IV).

Table 2 – Case selection strategies in the three variants of PT.

Within the two theory-centric variants of PT, typical cases are chosen that are members of X, Y and the relevant scope conditions for the mechanism. Case scores on other causal conditions are less relevant, given that we are focusing our attention on how a mechanism contributes to produce Y, not on the causal effects of an X upon values of Y.

Finally, deviant case selection strategies can be relevant in two situations. First and most widely used is when a mechanism has already been found between X and Y, but where there is a deviant case (zone II) where X has not led to Y. This type of case can be used to investigate omitted causal and/or scope conditions for the operation of the mechanism in question. Note that it is not only PT that enables us to do this; instead detecting these omitted conditions requires the use of comparative methods. Second, we can use theory-building PT to trace backwards from an outcome to detect a cause and mechanism. We argued, however, that comparative methods such as a MDSD offer more bang-for-the-buck in relation to this research purpose.

In conclusion, taking mechanisms seriously has major implications for case selection strategies in PT. Further, taking the differences across the three variants of PT seriously means that case selection strategies need to be appropriate for the variant chosen.

References

Allison, Graham and Zellikow Phillip. 1999. *Essence of Decision Explaining the Cuban Missile Crisis.* New York: Longman.

Beach, Derek and Rasmus Brun Pedersen. In Press. *Process-Tracing methods: Foundations and Guidelines.* Ann Arbor: University of Michigan Press.

Beach, Derek and Rasmus Brun Pedersen. What is process tracing actually tracing? The three variants of process tracing methods and the study of causal mechanisms. Paper prepared for presentation at The American Political Science Association annual meeting Seattle, Washington, USA, September 1-4, 2011.

Bennett, Andrew. 2008a. Process-Tracing: A Bayesian Perspective. In *The Oxford Handbook of Political Methodology,* ed. Janet M. Box-Steffensmeier, Henry E. Brady and David Collier, 702-721. Oxford: Oxford University Press.

Bennett, Andrew. 2008b. The mother of all "isms": Organizing political science around causal mechanisms. In *Revitalizing causality: realism about causality in philosophy and social science,* ed. Ruth Groff, 205-219. London: Routledge.

Braumoeller, Bear F. and Gary Goertz. 2000. The Methodology of Necessary Conditions. *American Journal of Political Science* 44(4): 844-858.

Checkel, Jeffrey T. 2008. Tracing Causal Mechanisms. *International Studies Review* 8(2): 362-370.

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

Day, Timothy and Harold Kincaid. 1994. Putting Inference to the Best Explanation in its Place. *Synthese* 98(2): 271-295.

Falleti, Tulia G. and Julia F. Lynch (2009) Context and Causal Mechanisms in Political Analysis. *Comparative Political Studies* 42:1143-1166

George, Alexander L. and Bennett Andrew. 2005 *Case studies and Theory Development in the Social Sciences* Cambridge Massachusetts London: MIT Press.

Gerring, John. 2007. *Case Study Research.* Cambridge: Cambridge University Press.

Gerring, John. 2010. Causal Mechanisms: Yes But... *Comparative Political Studies* 43(11): 1499-1526.

Gerring, John. 2012. Mere Description. *British Journal of Political Science* First View, July 2012: 1-26.

Glennan, Stuart S. 1996. Mechanisms and the Nature of Causation. *Erkenntnis* 44(1): 49-71.

Goertz, Gary. 2006. *Social Science Concepts: A User's Guide.* Princeton: Princeton University Press.

Goertz, Gary. 2008. Choosing Cases for Case Studies: A Qualitative Logic. *Qualitative and Multi-Method Research* Fall 2008: 11-14.

Humphreys, Adam RC. 2010. The heuristic explanation of explanatory theories in International Relations. *European Journal of International Relations* 17(2): 257-277.

Jackson, Patrick T. 2011. *The Conduct of Inquiry in International Relations.* London: Routledge.

Janis, Irving L. 1983. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes.* Boston: Houghton Mifflin Company.

Jervis, Robert. 2010. *Why Intelligence Fails: Lessons from the Iranian Revolution and the Iraq War.* Ithica: Cornell University Press.

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

Layne, Christopher. 2006. *The Peace of Illusions: American Grand Strategy from 1940 to the Present.* Ithica: Cornell University Press.

Levy, Jack. 2002. Qualitative Methods in International Relations. In *Evaluating Methodology in International Studies,* ed. Frank P. Harvey and Michael Brecher, 131-160. Ann Arbor: University of Michigan Press.

Lieberman, Evan S. 2005. Nested Analysis as a Mixed-Method Strategy for Comparative Research. *American Political Science Review* 99(3): 435-451.

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

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

Mackie, J.L. 1965. Causes and Conditions. *American Philosophical Quarterly* 2(2): 245-264.

Mahoney, James and Gary Goertz. 2004. The Possibility Principle: Choosing Negative Cases in Comparative Research. *American Political Science Review* 98(4): 653-669.

Rohlfing, Ingo. 2008. What You See and What You Get: Pitfalls and Principles of Nested Analysis in Comparative Research. *Comparative Political Studies* 41(11): 1492-1514.

Ragin, Charles C. 2000. *Fuzzy-Set Social Science.* Chicago: University of Chicago Press.

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

Ross, Michael L. 2001. Does Oil Hinder Democracy? *World Politics* 53(2): 325-361.

Schiff, Benjamin N.. 2008. *Building the International Criminal Court.* Cambridge: Cambridge University Press.

Seawright, Jason. 2002. Testing for Necessary and/or Sufficient Causation: Which Cases Are Relevant?. *Political Analysis* 10(2): 178-193.

Sil, Rudra and Peter J. Katzenstein. 2010. *Beyond Paradigms: Analytical Eclecticism in the Study of World Politics.* Basingstoke: Palgrave Macmillan.

Van Evera, Stephen 1997. *Guide to Methods for Students of Political Science.* Ithica: Cornell University Press.