

Going beyond the single case: Comparative Process Tracing as a tool to enable generalizations about causal processes

Paper prepared for the 2022 Annual Meeting of the American Political Science Association, Montreal, Canada.

Authors:

Derek Beach, Aarhus University (derek@ps.au.dk)

Gabriela Camacho, Aarhus University (gabriela.camacho@ps.au.dk)

Markus Siewert, TUM Think Tank (markus.siewert@hfp.tum.de)

Abstract

Process tracing methods have been considerably developed in the past decade, with most developed focused on how to make sound internally valid inferences about causal processes. There has been less attention given to the external validity of processual claims. As a result, most PT studies are one-off studies that do not allow for cumulative knowledge building about causal processes outside of the original case. Another strategy is to make sweeping conclusions from case studies using one-to-many generalizations. However, this presumes that cross-case causal homogeneity implies within-case homogeneity, without engaging in an empirical analysis of whether this is the case. In this article, we argue that this is problematic for three reasons: 1) omissions, 2) lumping, and 3) sequences and interactions.

We then develop a comparative process-tracing strategy to validate empirically the scope of cases to which process-level generalizations can be made. This strategy has two phases, intensive and extensive. The intensive phase involves comparing results from two or more PT case studies, both in terms of whether each part of the process was similar and whether the conditions were similar. If a different process is found, then the analysis should focus on identifying the difference between the cases that accounts for different processes. If the causal process operating in the two initial cases is similar, we move to the extensive phase, where the researcher conducts PT 'light' studies of ever more diverse cases to explore empirically the bounds of valid processual generalizations.

Section 1 - Introduction

In this article, we put forward a step-by-step procedure for using comparative PT in order to increase the external validity of processual claims.¹ Comparative PT provides scholars with an analytical procedure for empirically exploring the types of cases within which we can expect a given causal process to work, enabling evidence-based processual generalizations to be made. The goal is to produce *evidence-based* processual generalizations that shed light on the conditions under which we should expect a given process (or processes) to link a cause (or set of causes) and an outcome.

In the past decade, process tracing (PT) methods have been considerably advanced. Most of the developments have focused on how to make internally valid inferences about causal processes (Bennet and Checkel, 2014; Waldner, 2014, 2015; Mahoney, 2015; Beach and Pedersen, 2016, 2019; Fairfield and Charman, 2017). Much less attention has been given to the question of the external validity of processual claims. It is therefore not surprising that most published PT case studies produce one-off analyses that either side-step the question of whether similar processes might be operative in other cases, suggest that it might be present in other cases but without any evidence, or engage in a very short analysis of one or two other cases to document that the process can be found in other cases. As a result, there is typically little cumulative knowledge-building about how things work in different contexts as a result of published PT case studies.

Another common strategy is to use case selection from a population of cases to make a one-to-many generalization.² Here a case is selected based on cross-case analysis that is found to be typical/representative or least-likely in relation to a given population. After finding evidence of a process within the selected case, the study will then claim that similar processes should be present in the population. However, making a one-to-many

¹ - While we treat the term causal process as a synonym to the term causal mechanism, we do not use the later term because it is understood to mean so many different things in the literature. The term causal process is more straightforward, and is used to refer to what links a cause and outcome together in a causal relationship, i.e. the linkage between the two.

² - Less common are scholars who engage in an 'all-cases' strategy, meaning that they use some form of case study analysis to assess all of the cases in the population, although typically not using in-depth PT (e.g. Haggard and Kaufman, 2016). Here there are of course no generalizations being made, meaning that the external validity issue is irrelevant.

generalizing inference builds on strong assumptions about causal homogeneity in the target population. Even more problematic, even if we have some form of cross-case analysis (e.g. regression-based or QCA) that suggests that the target population is relatively causal homogeneous at the level of causes and outcomes, we would still be in the dark about whether similar *processes* are operative within other cases. Therefore, a one-to-many generalization is a form of ‘one-size-fits-all’ claim that lacks empirical evidence. As Donal Khosrowi (2019:48) has put it, ‘[...] merely assuming that populations are similar at lower levels would amount to extrapolation based on hope...’. He goes on to argue that instead, scholars ‘...need to support empirically the claim that populations are sufficiently similar to warrant extrapolation.’ (ibid).

Comparative PT has the ambition of enabling this type of evidence-based processual generalizations to be made. They will typically be more contextualized claims to bounded target population of cases.³ This article develops comparative PT as a method for making this type of evidence-based process generalizations. The article proceeds in three steps. In section 2 we present the core challenge of causal homogeneity at the level of causes/outcomes and causal processes. In section 3 we unpack why causal heterogeneity at the process-level can be masked by what appears to be homogeneity at the level of causes and outcome, which means we cannot assume homogeneity at both levels. In section 4 we develop a methodological strategy for engaging in comparative PT that is focused on increasing the external validity of processual claims. Our procedure involves engaging in relatively intensive PT analysis of at least two cases, which are compared at both the cause/outcome and process level. The intensive analysis is then followed by a more extensive analysis that uses a form of PT ‘light’ to search for potential processual heterogeneity across a strategic sample of ever more diverse cases. The extensive analysis can be supplemented at later stages by analysis of deviant (consistency) cases, in which the process should have linked the cause and outcome together but where it broke down (Beach and Pedersen, 2019: 274-5). The deviant cases are assessed to shed more light on the conditions required for a process to operate by both tracing the process until it breaks down, and then comparing it with one or more typical cases to figure out what was missing in the deviant case.

³ - See Bechtel and Richardson, 2010 for similar arguments from the natural sciences.

Section 2. The problem of potential processual heterogeneity

The methodological challenge tackled by this article is how to produce empirical evidence that can increase our confidence about the conditions under which we should expect a given causal process to operate.⁴ In this section, we unpack why we cannot merely assume that causal homogeneity at the level of causes/outcome implies homogeneity at the level of causal processes.

However, before we can proceed it is important to define what we mean by causal process (aka causal mechanism). While the nature of causal processes is heavily contested in the literature, at a minimum most scholars agree that theories of causal processes explain what *links* causes and outcomes together in some form of dependency or productive relationship (i.e. causal relationship) (see a review in Beach, 2022). This also means that studying *how* a causal process works is per definition a form of *within-case* analysis; causes are causally linked with outcomes *within* cases, not across cases (Beach and Kaas, 2020). In contrast, assessing whether X has a causal effect on Y using a variance-based design (e.g. an experiment), or whether X has an invariant association with Y using QCA (case-based) involves cross-case analysis of patterns of variance/invariance across cases. Given that causality is studied across cases in these designs, the analysis does not shed light on causal relationships in any given case – which is why some scholars term the type of causal inferences made using cross-case analyses as ‘generic causality’ (e.g. Cartwright, 2011; Clarke et al. 2014). This difference is seen in figure 1, below. On the left-side, cross case analysis using set-theoretical comparative methods such as QCA would find that C1 and C2 in conjunction were (almost)⁵ sufficient for the outcome (O) to occur. However, the cross-case analysis tells us nothing about what links C1 and C2 with the outcome in any given case. Similarly, on the left-side is a controlled experiment, where one group receives X_t (treatment, i.e. cause), and the other receives a control (X_c). The cross-case analysis investigates the difference on average between values of Y in the treated and controlled population. But the cross-case analysis also tells us nothing about what links

⁴ - Note that our arguments are not isolated to either a case-based or variance-based approach, but unless otherwise stated, apply to both. In a case-based approach, terms like conditions and outcomes are used, whereas in a variance-based approach the term variable is used.

⁵ - Here there is one case where C1 and C2 are present but O is not, i.e. it is not consistent with $C1 * C2 \rightarrow O$. Most QCA analysts are satisfied with finding an invariant pattern that is not 100% consistent.

X and Y in any given case. To do so requires within-case analysis – for example in the form of PT.

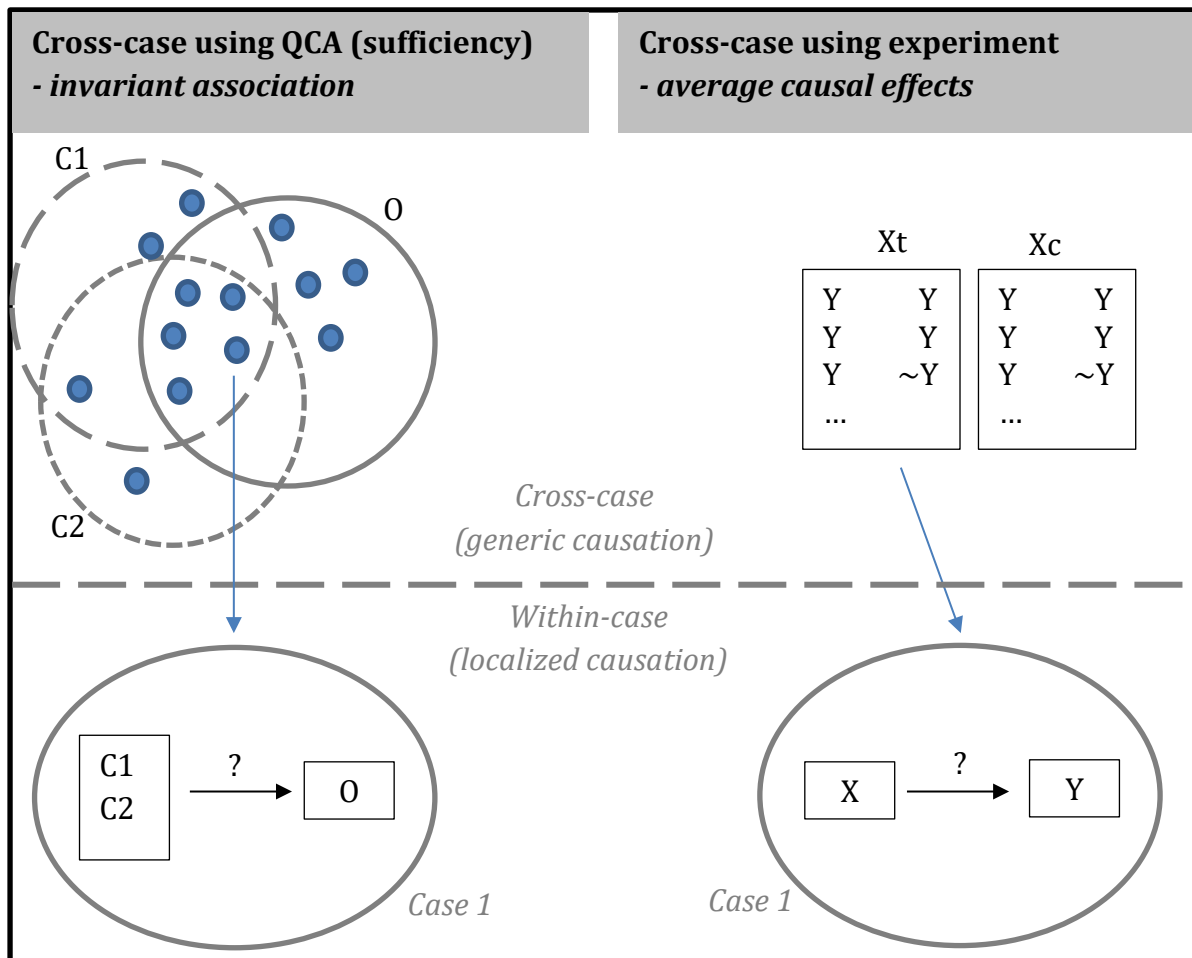


Figure 1 – cross-case and within-case analysis.

Causal homogeneity at the level of causes/outcomes means that within a given set of cases, the same cause will have the same effect on the outcome [i.e. the Stable Unit Treatment Value Assumption, SUTVA (Morgan and Winship, 2007: 37-40)], whereas heterogeneity means that the same cause can have different effects depending on the presence of other conditions/covariates. Causal homogeneity at the level of causal processes linking X and Y together means that the same X is linked to the same Y through similar causal processes (aka causal mechanisms)⁶ (Beach and Pedersen 2016; 2019; Beach and Rohlfing 2018; George and Bennett 2015; Gerring 2010; Goertz 2017; Salmon 1998; Sayer 2000; Steel 2008; Weller and Barnes 2016). Processual heterogeneity refers

⁶ - Note that they do not have to be exactly similar. See below for more discussion about varying levels of abstraction of processual theories.

to the situation where different processes link the same cause/outcome together under different conditions. In other words, the same X/Y relationship can be linked through equifinal causal processes.⁷ This can mean either that the whole process is different, or only parts differ across cases. In some circumstances, X might trigger one process, whereas in another context X might trigger multiple processes that are linked with the same outcome.

The challenge is that PT only enables inferences about causal processes that are operative in the studied case. Given the analytical resources required for good PT, and in particular the need to actually trace empirically the workings of each of the critical parts of a process, it is usually impossible to undertake more than one or two in-depth PT studies in any given research project. Article word lengths make it even more difficult, and typically restrain the scholar from doing more than one in-depth PT, followed by one or more superficial PT 'light' analyses (see section 4 for more on 'light' versions of PT). As a result, many published PT studies are one-offs, and there is little cumulative knowledge about *how* things work.

The existing alternative to this is to merely *assume* that what one found in the studied case(s) should be found in other cases that look similar at the cross-case level (e.g. Gerring, 2007). Relying on the assumption that causal homogeneity at the level of causes/outcomes implies processual homogeneity naturally makes life easier as regards processual generalizations. For instance, in Schneider and Rohlfing's accounts of the combination of PT and cross-case analysis using QCA, they suggest that in the absence of model misspecification, '...the same sufficient term cannot give rise to different mechanisms' (Schneider and Rohlfing 2016:555). Assuming that there is not model misspecification then enables a one-to-many processual generalization, where the

⁷ - Processual heterogeneity can also refer to situations of multifinality and masking. With multifinality, the same causes in differing contexts trigger the same processes that lead to different outcomes (for instance, Falletti and Lynch 2009:1151). However, QCA or statistical analysis would detect this situation, which is why we focus on equifinal causal processes in this article. Masking instead refers to the situation where a cause triggers counter-directed processes that can have different effects on the same outcome (Gerring 2010:1511; Steel 2008:68). The classic example here is the relationship between exercise and weight. Exercise triggers several causal processes, including a 'calorie burning' process that decreases weight, and a 'muscle-building' process that increases weight. Both processes would be active in any given case, but depending on a range of factors such as type of exercise, duration and diet, exercise will increase or decrease weight.

'[f]indings from the study of one, say, typical case, travel to all other typical cases of the same term, but not beyond this term' (ibid.:556). In terms of figure 1, after finding a causal process (CP) in case 1, we would then assume that a similar CP linked C1 and C2 with the outcome in the other cases that are members of the term $C1 * C2 \rightarrow O$.

However, assuming processual homogeneity based on causal homogeneity at the level of causes/outcome is problematic given that, for all but the simplest causal processes, how they play out can vary considerably for reasons to be elaborated in section 3, below (Salmon 1998; Bardach 2004; Steel 2008; Falletti and Lynch 2009; Gerring 2010; Xie et al 2012; Craver and Darden 2013; Clarke et al. 2014; Weller and Barnes, 2014: 21; Morgan and Winship 2015:48–52; Khosrowi 2019). Against this backdrop, the problem of generalizing processual claims from one studied case to other cases is a crucial one in the social sciences, and beyond, but one that we still do not have good methodological tools for tackling (Cartwright, 2011; Khosrowi, 2019; Steel, 2008; Wilde and Parkkinen, 2019).

This is not just an abstract theoretical problem, but is a real-world issue which characterizes most research situations in the social sciences. One prominent example of processual heterogeneity can be found in the study by Ross (2004) on the connection between the presence of natural resource wealth (cause) and the onset of civil war (outcome). In his work, he suggests six different quite abstract processes, which are not necessarily mutually exclusive to each other, meaning that there can be one or more present at the same time in a given case (ibid.:39-42, 57; see also Figure 2a). The first processes relate to capacity building processes, whereby potential rebels either loot/extort primary products (M1) or sell future contracts to outsiders (M2), enabling them to purchase weapons and recruit soldiers to start a rebellion. The third and fourth processes deal with types of grievances, where either the pollution and corruption created by resource extraction creates grievances amongst local populations (M3), or centralized control of production produces incentives for local elites to support rebels in order to get a piece of the action leading to separatism (M4). The fifth process relates to price instability in primary product markets which prevents the government from both spending as much as it might want on the coercive apparatus and produces lower social spending (M5). Finally, neighboring powers might want to intervene in a country by supporting rebels in order to gain wealth from the primary product production (M6).



Figure 2a: Processual heterogeneity as hypothesized by Ross (2004)

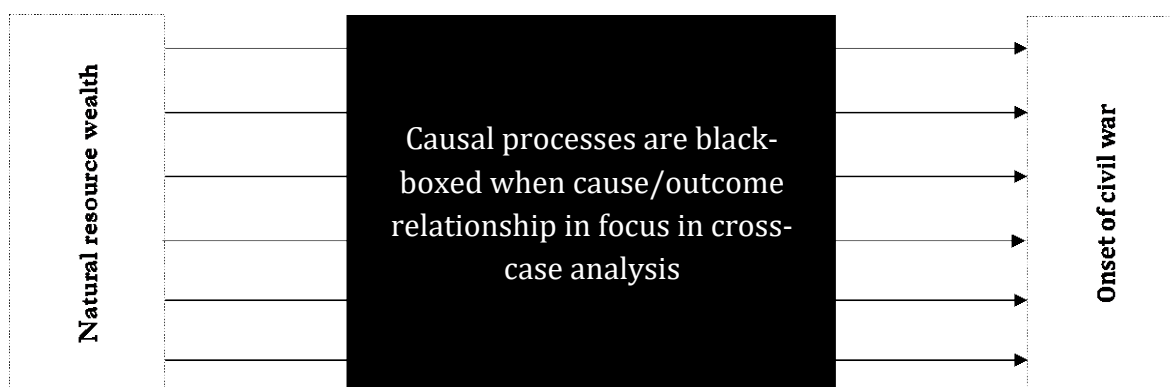


Figure 2b: What a cross-case analysis (e.g. QCA) tells us about which processes are at play

Figure 2 – Processual heterogeneity and the link between natural resource wealth and onset of civil war

Source: Own depiction based on Ross 2004:39-42.

A cross-case analysis (e.g. with QCA) might which have revealed that natural resource wealth is sufficient for the onset of civil war, but it would tell us nothing about what processes linked cause/outcome together in any given case. If we did a case study and found M1 in one of the positive cases of wealth/civil war, we would however not be able to generalize that it is the same processes in other cases given the number of different potential processes that could be at play in any given case. It is here that comparative PT methods are relevant as a tool to ensure the external validity of processual claims.

Section 3 - Sources of processual heterogeneity

In this section we unpack three distinct reasons why it is problematic to assume that causal homogeneity found using cross-case methods means that there is homogeneity in how X and Y are linked through causal processes. The three reasons are: 1) omissions, 2) lumping, and 3) sequences and interactions. In the final part of this section, we address the critique that processual heterogeneity is ever-present, showing that merely lifting the level of theoretical abstraction of processual claims is not a solution.

3.1. omission of conditions/variables

The omission of conditions/variables is typically what we think of when talking about contextual conditions (scope) (e.g. Falletti and Lynch, 2009). Omissions can be 'known', where conditions/variables are dropped from the causal model because the cross-case analysis has found that a condition/variable either do not produce a difference in the outcome (lack of variance), or when cross-case analysis is done using QCA (sufficiency), there is a difference in the condition but similar outcome (lack of invariance) that would lead to the condition being found redundant/irrelevant. However, just because our cross-case analysis tells us that a given condition/variable does not matter, we cannot automatically assume that the condition/variable does not matter at the processual level.

As an example of where it would be problematic to assume homogeneity when conditions are dropped from the model, in a study by Mello (2012) on military (non-)participation in the Iraq war. Here, the minimalization of the model done in the QCA analysis revealed that countries deployed troops if a conservative executive (E) was unrestricted by parliamentary veto rights ($\sim V$) and free of constitutional provisions on their capacity to conduct war ($\sim C$) (thus, $E^* \sim V^* \sim C$; *ibid*:437-442). This combination was found to be sufficient, irrespective of whether the ideological center in parliament leaned to the right, like in Australia, the UK or the US (condition P), or the left, like in Spain, Poland or Portugal (condition $\sim P$). If we followed a one-to-many strategy, the best choice for doing a PT case study would be Australia (which is the most typical case) to reconstruct the underlying causal mechanisms and explain how the decision to deploy troops played out (Schneider and Rohlfing, 2016). The processual claims might then be generalized to all other cases included in the final term $E^* \sim V^* \sim C$, irrespective of P or $\sim P$. However, a

generalization strategy which is sensitive to processual heterogeneity would be more cautious and explore empirically whether the processes differ depending on where the ideological center is (P/~P). It is not difficult to imagine that the negotiation processes between the executive and parliament might differ in the two situations: one where an unrestricted, right executive is backed by a right-leaning parliamentary majority (Australia), and in the other a conservative government faces a left-leaning ideological center in parliament (Spain).

Unknown omissions deal with conditions/variables that either were not considered for the original causal model, or they were not included because there were not expected to be causally relevant at the cause/outcome level. However, this type of omission might impact how things play out at the process level. One particularly acute omission is when there is the shadow of the past that can lead actors to behave differently in relation to similar circumstances. For instance, if there had been a previous coup in a country, it might impact future relations between political and military actors who both attempt to learn from history by acting in a different manner. If we were to compare this case with another country in which there was no legacy, actors might react quite differently to similar stimuli across the two cases. Here the shadow of the past would be an 'unknown' omission that might produce different processes.

As an example of an unknown omission that might impact processes that was not included in the original cross-case analysis, Bretthauer (2015) examined the conditions that contribute to intra-state conflicts in situations of resource scarcity, including lack of fresh water or arable land. In the analysis of negative cases, i.e. countries which do not experience civil war, a wide set of cases that are found to be causally homogeneous at the level of conditions/outcomes. However, the set of negative cases includes very disparate countries, with Cape Verde, Kyrgyzstan, Singapore, Bhutan, and Arab Emirates together with Chile, Japan, the Netherlands, or Switzerland. However, it is not difficult to imagine that different processes might be operative in Switzerland than in Singapore, Cape Verde due to conditions that were not included in the original model (ibid.:606-609).

3.2. lumping cases together when conceptualizing

When scholars define concepts, they are often thinking about defining concepts in a way that ensures that categorization of cases results in (relatively) causally homogeneous cases (Goertz, 2006). The scale used to measure concepts matters. Differences-in-kind are used to capture causal differences (e.g. presence of X produces Y, whereas absence of X produces Z), whereas differences-in-degree capture gradual increases/decreases in the strength of the causal relationship. Additionally, the attributes included in a concept and how they are combined also matters. It is problematic to assume that cases that are lumped together as causally similar during conceptualization are also causally similar at the processual level.

Concepts come in various forms as discussed in the literature on concept formation (Barrenechea and Castillo 2019; Collier and Levitsky 1997; Goertz 2006; Møller and Skaaning 2010).⁸ If the conceptualization of a cause (or set of causes) or outcome contains multiple attributes, each attribute might have its own causal properties and thus potentially trigger its own, different causal process(s). Yet cross-case methods only use the *aggregate* score for a specific concept, with the result that multiple equifinal processes might lurk underneath the same concept/measure. A very intuitive example for how complex concepts might be linked to processual heterogeneity can be illustrated in concepts defined in family resemblance terms. By definition, concepts following a standard family resemblance (OR) structure contain multiple attributes which are each individually sufficient for the presence of a concept. The assumption behind a family resemble conceptualization is that the different attributes are functionally equivalent in relation to effects on the outcome (Goertz, 2006: 32-46). However, because two or more attributes show the same effect concerning their outcome, this does not mean that the same processes link these different attributes to the same outcome.⁹

⁸ - Conceptual attributes are frequently not systematically spelt out or it might unclear what the actual 'causal' properties are expected to be. For instance, when using an aggregate measure of democracy, it is often not clear what components are included and what mechanisms are conceived as possible. For a related argument, see Rutten 2020:12-13.

⁹ - Barrenechea and Castillo (2019:125-27) point out that family resemblance concepts are heterogenous by design. They use the example of a family resemblance concept as dependent variable where different pathways might be linked to different configurations (aka types of cases) of the outcome.

How concept structure complicates processual generalization can be illustrated using Binder's (2015) QCA study of robust UN interventions in international conflicts. Here, the condition labeled 'spillover effects' of a given conflict to neighboring countries plays a crucial role. Spillover effects are conceptualized via three attributes: i) refugee flows, ii) transnationally operating rebel groups, and iii) other negative effects such as drug traffic or economic downturns which are combined into a single set score based on a family resemblance approach (ibid.:715-16). Yet it is not difficult to envisage that very different causal processes might be triggered depending on which of these attributes is present in a given case. In some cases, only one attribute might be present (refugee flows or transnationally operating rebels or economic downturns), whereas in other cases a combination of two or even all three might be present. If we then studied one positive case, it would be very problematic to generalize about processes in other cases because there might be very different attributes present in other cases. In this situation, it would be problematic to assume similar processes in other positive cases – at best, we are able to potentially generalize to cases which share the same configuration of attributes for the condition.

The scale used to define a concept can also produce processual heterogeneity. As an example, in an article by Kuehn and Trinkunas (2017), the focus is on the conditions that produce military contestation in a set of Latin American countries in the 1990s and 2000s. Three conditions are found to produce the outcome: leftist populist ideology of the leader, degree of radicalism of the leader, and the presence of resource rents. Resource rents is defined as the share of rents / GDP at the start of the case period using data from the World Bank. The concept is defined in fuzzy-set terms, in which the Chavez 2 case (2001-06) and the Correa 1 case (2006-09) are both in the set, and are differentiated only by differences-in-degree (scores of 0.98 and 0.7 respectively). However, when we assess the condition resource rents and how cases are coded in light of potential processual linkages, we find that resource rents in the Chavez 2 case were steadily increasing during the period (from 13,2% of GDP in 2001 to 30.1% in 2006). In this case, the processes triggered involved conflicts between the leftist, radical populism of Chavez and the military that were fueled by the oil bonanza. If we look at the cross-case analysis (QCA), other cases like the Correa 1 case in Ecuador are part of the same solution set, meaning that in theory we could then generalize the process found in the

Chavez 2 to the Correa 1 case. However, if we look at Ecuador during this period, there was a massive *decrease* in oil rents (from 18.4% in 2007 to 8.8% in 2008), which created the need for the austerity policies that triggered protests from the police that were coded as military contestation by the authors. Given that different processes can plausibly be triggered by the cases that were lumped together due to how resource rents was conceptualized, it would be highly problematic to assume process homogeneity based on the cross-case results.

3.3. temporal and interactive dynamics

A final reason for why causal homogeneity at the cause/outcome level does not necessarily imply processual homogeneity is that cross-case analysis *typically* is not sensitive to whether sequences matter (e.g. is X1 before X2 causally different than X2 before X1), or how interactions matter (e.g. X1 has a causal effect on Y, whereas when X2 is present the effect disappears). In variance-based approaches there are many statistical tools to probe sequencing and interactions and whether they matter for variation in Y. However, this form of cross-case analysis does not take into consideration whether differences in sequence or interactions might impact which processes are triggered.

In case-based comparative methods (QCA), while being able to identify conjunctions is an important feature, configurations only show that two or more conditions co-occur together. Configurations tell us nothing about the dynamics and interactions *between* the conditions which constitute the configuration, let alone about sequencing or other temporal aspects (Baumgartner 2013; Beach and Rohlfing 2018). In other words, the dynamic character of the constituting conditions is black-boxed by QCA, hence concealing potential processual heterogeneity which needs to be tested for. We suggest that one should explore whether similar mechanisms are present in different cases *within* a given solution set. Note that techniques like temporal QCA (Ragin and Strand 2008), or Coincidence Analysis (Baumgartner 2013) are also not able to unpack the interplay among conditions at the level of mechanisms.

3.4. Why abstraction of process theories is not the solution

At this point, the skeptical reader might point out that the problem of processual heterogeneity is mainly a matter of *analytical abstraction*. Indeed, mechanisms can be

conceived as very case-specific or as broad and relatively abstract phenomena. Logically, the more abstract a theorized mechanism (lower intension), the more cases can be found in where it is present (higher extension), and vice versa.¹⁰ This is depicted in the columns of Figure 3, below.

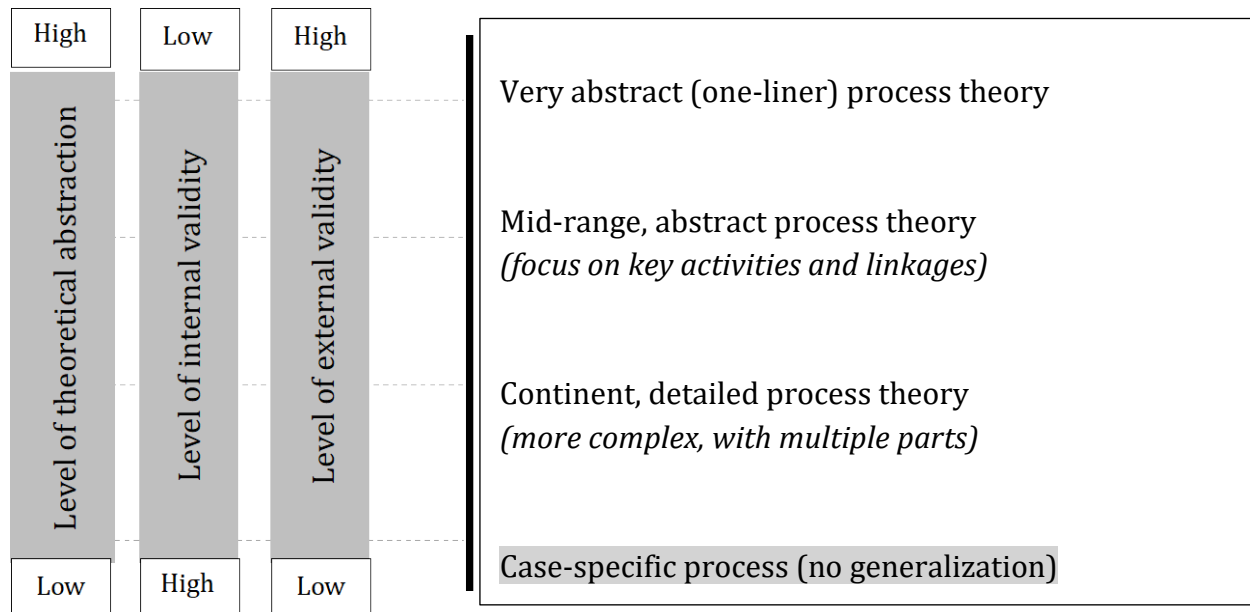


Figure 3 – The relationship between theoretical abstraction and internal/external validity of processual claims

Source: Own depiction.

Some scholars suggest that processual theories should be so abstract that they are literally context-free. For instance, when discussing causal processes (aka mechanisms), Jon Elster (1998:52–55) uses abstract terms like ‘wishful thinking’ or ‘the spillover effect’. However, what is important to note with Elster’s usage of the term causal process is that the theories he depicts are so abstract that besides being context-free, they are also cause and outcome free! The question of what causes a ‘spillover effect’ is not answered in his work, nor is what outcome its operation could produce. As a result, we do not learn about ‘how it works’ which is the key value-added of process-focused research (Craver and Darden, 2013; Clarke et al, 2014).

¹⁰ - There can be exceptions to this rule as Craver and Kaplan (2020:307) correctly note that ‘decreasing detail (increasing the degree of abstraction) of a model does not always increase its scope’. For instance, we might have a very abstract, rough sketch of a mechanism which only pertains to a handful of cases.

Another critique would be that we should not understand process theories as representatives of causal structures that might play out in similar ways across different cases. Instead, we should think of process theories as ideal types that can be used as heuristic tools to understand how a *particular* case deviates or reflects the ideal typical theory (Saylor, 2002:1002-5). An ideal type is then an, ‘...analytical construct [that] is found nowhere in empirical reality; it is a utopia.’ (Weber, quoted in Jackson, 2016: 154). However, this means that we move away from the goal of trying to understand similarities and differences in how causal processes play out in different types of cases. Ideal types can per definition are focused on understanding the particular.

If we accept that causal processes are what link causes and outcomes together, and we have the ambition of trying to move beyond understanding how things work in particular cases, the issue of processual heterogeneity rears its head again. This is true even if we are operating with quite abstract one-liner type process theories similar to what Ross theorized (see above in section 2). The problem obviously is more severe when we lower the level of abstraction of our process theory¹¹ by unpacking more of the ‘blow-by-blow’ steps of the process linking causes and outcomes.

Returning to the Ross (2004) example, a highly abstract theorized mechanism linking resources with civil war might be ‘rebels loot oil production’, which obviously tells us little about what rebels and other social actors are actually doing in a case. Here the processual theory only ‘...describes some of the internal details of the mechanism [process] but has black boxes signifying that one or more relevant component parts, activities, and organizational features are unknown’ (Craver and Kaplan 2020:299). Given this black-boxing, we would not be able to make a strong causal inference about the linkage between resources and civil war because important parts of the causal story are not evidenced. Most obviously, just because the rebels were looting oil production

¹¹ - - The level of theoretical abstraction should not be confused with moving to the evidential level of a case. When we lower our level of theoretical abstraction, we are unpacking how the process works in terms of theorized interactions between social actors. In the example below, evidence of the abstract process of ‘rebels looting’ would be case-specific (e.g. press accounts of a particular rebel group stealing oil in one case, whereas we might have testimonials from government officials in another). But the evidence of the operation of a process, or parts thereof, will always be case-specific.

does not mean that this was actually linked in a continual process with the outbreak of civil war. In other words, the internal validity of processual claims would be quite low.

In order to make stronger causal inferences about the process, we would want to lower the level of theoretical abstraction to enable us to trace the critical parts of the process empirically,¹² e.g. by not only looking at ‘rebels looting’, but also whether they used the money to buy weapons, and whether they then used these weapons to stage a conflict with the government. This formulation is still a relatively abstract—a ‘mid-range’ process theory - which is abstract enough to potentially be operative across many cases (Cartwright, 2020). Causal inferences would then be possible by basically ‘following the money’, trying to find out whether there is evidence of the sequential operation of each part of this process. However, because of the relatively high level of abstraction, there would still be holes in the causal story, meaning that we cannot make very strong causal inferences because we have not evidenced an unbroken chain between the cause(s) and outcome.

If our research goal is to make strong causal inferences and/or understand how the process played out in a *particular* case, we would need to provide ‘[...] all of the entities, properties, activities, and organizational features that are *relevant* to every aspect of the phenomenon to be explained’ (Craver 2006, 360 [italics inserted]; see also, Craver and Kaplan 2020).¹³ This would mean we might even descend to the level of case-specific, detailed mechanistic theorizations. For example, if we were trying to understand the role resources played in the outbreak of civil war in Syria, relevant aspects would include unpacking how a rebel group such as ISIS was able to steal oil and sell it on the black market that existed in the chaotic conditions in the region, and were then able to purchase weapons that were used in the civil war. Getting to this level of theoretical detail would

¹² - Another benefit of lowering the level of processual abstraction is that when theorizing activities and linkages and studying them empirically, we also shed light on the supporting conditions that might be relevant for where a process operates. We return to this point in section 4.

¹³ - The term ‘relevant’ is important here because it disciplines how much further we go into the gory details of cases, with the pragmatic standard of when to stop being when further details do not contribute further explanatory leverage in relation to the purpose the mechanistic theory plays in our research. There are, however, no clear standards nor an automatic formula but multiple metrics to assess the relevance of mechanistic details. What counts as relevant therefore always needs to be answered with respect to the actual research at hand which might differ depending on the theoretical perspective, the current state of the art (i.e. what do we already know about the working of the mechanism), among others (Craver and Kaplan 2020, 305, 310–13; see also various contributions in Glennan and Illari 2017).

enable relatively strong causal inferences because each part of the process would be evidenced by tracing them empirically (naturally contingent on the quality of the empirical material available).

However, most scholars engaging in PT analysis have both the ambition to make *relatively* strong causal inferences while also being able to say something about processes beyond the scope of the studied cases, meaning the goal is to land somewhere between the extremes of very abstract and case-specific processual theories; i.e. the realm of **mid-range process theories**. Mid-range process theories involve significant analytical abstraction, with the focus on the critical parts of the causal 'story', and in particular the activities of actors and their interactions. We define critical phases as the parts of the process that are particularly crucial from a causal perspective, and where we have theoretical reasons to expect that the process might most plausibly differ across cases (Steel, 2008: 88-92). For example, if we are studying a policy learning process, we would expect that some form of lesson-drawing from the past phase of the process would take place – otherwise there would be no actual learning taking place.

The activities and actors in a mid-range process theory are theorized in quite abstract terms, and the theory typically will only have three to five parts. A good example of a mid-range process theory can be found in O'Mahoney (2017), where he posits a 'rhetorical adduction' process theory in quite abstract terms. This is depicted in table 1, below. In the process theory, the cause (or trigger) is two groups of states that have a dispute, and the outcome is a new policy being adopted. The article applies the process theory to understand the actions taken by the parties in the Indo-Pakistan war of 1971, and how the shifting of the debate by India (claimant) through its actions on the ground vis-a-vis Pakistan (opposition) led to the subsequent recognition of Bangladesh despite initial resistance from the international community (audience).

<i>Cause (trigger)</i>	<i>Rhetorical adduction causal process</i>				<i>Outcome</i>
	<u>Part 1</u> (Object)	<u>Part 2</u> (Frame)	<u>Part 3</u> (Adduce)	<u>Part 4</u> (React)	
Two group of states with dispute over policy (enact or not) →	<u>Opposition</u> argues that policy is illegitimate because has property X →	<u>Claimant</u> argues that policy is legitimate because has property Y, not X →	<u>Claimant</u> takes action that links policy with Y →	<u>Audience of undecided states</u> acquiesces/ supports policy because made more plausible Y than X →	Policy adopted

Table 1 – Mid-range causal process linking policy dispute with new policy adoption.

Source: based on O'Mahoney, 2017.

Note that abstract terms are used by the author (e.g. claimant state), and the step-by-step process focuses only on the most critical steps and linkages without theorizing everything that actors are doing in their interactions with each other. Further, the activities themselves are quite abstract, meaning that they might lump together different activities that are functionally equivalent to each other. For instance, arguing might involve quite different activities in actual cases, but where they are functionally equivalent (i.e. how they work in a causal sense is similar across the different particular activities).

Comparative PT should compare mid-range process theories and whether there is process evidence for similar actors engaging in similar activities for each of the critical parts of the process. This is depicted in table 2, below.

<i>Cause (trigger)</i>	<i><u>Comparison – similar process at critical phases?</u></i>			<i>Outcome</i>
	<i>Part 1</i>	<i>Part 2</i>	<i>Part 3</i>	
<i>Cause</i>	<i>actor engaging in activity</i>	<i>actor engaging in activity</i>	<i>actor engaging in activity</i>	<i>Outcome</i>
<i>Cause</i>	<i>similar actor engaging in similar activity?</i>	<i>similar actor engaging in similar activity?</i>	<i>similar actor engaging in similar activity?</i>	<i>Outcome</i>

Table 2 – Comparisons at the level of mid-range causal processes.

When can we claim that mid-range causal process theories are sufficiently similar to be able to infer that it is actually a similar process? Remember that a mid-range process theory abstracts away the case-specific details of cases. Unfortunately, there is (probably) not a strict standard to which we can refer when determining whether processes are sufficiently similar. We provide a longer example of what mid-range similarity and differences can look like in the next section.

Pragmatically, the criteria to infer similarity at the level of mid-range process include:

- did the parts take place in a similar sequence? Were there important parts of the process that were not similar?
- were the actors similar in causal terms (e.g. a policy entrepreneur might be from an advocacy NGO in one case, and from a think tank in another, but in relation to the process what only mattered was that a non-state actor tabled a policy proposal)?
- were the activities similar, meaning did they provide a similar causal linkage to the next part of the process (i.e. were the activities functionally equivalent)?

In the next section, we detail how this form of intensive comparative PT can be undertaken. We also contend that for analytical resource reasons, extensive comparative PT should be used as a follow-up, in which ever more diverse cases are assessed at the abstract process level.

Section 4 – Comparative PT

In this section we put forward a step-by-step procedure for comparative PT that can enable evidence-based process generalization to be made. The comparative element involves comparing results from two or more PT case studies, both in terms of whether each part of the process was similar, and whether conditions/variables were similar. When a *different* process is found when comparing processes in two otherwise *similar* cases, the comparison should then focus on *identifying the difference between the cases* that can explain the *different* processes.

We distinguish between what we term an ‘intensive’ phase that involves comparisons of two (or more) in-depth PT case studies, and an ‘extensive’ phase that involves successive iterations of more superficial PT case studies to explore the bounds of valid processual generalizations. Figure 5 illustrates the two phases and how it enables either evidence-based process generalizations to be made to a broader target population, or if processes differ, demarcation of the population into sub-sets, within which there is processual homogeneity.

Note that how each of the phases is implemented depends on the amount of resources possessed, but also due to the causal complexity of the phenomena being studied. In this respect, comparative PT is not a one-size-fits-all mechanical procedure, but should be tailor-fit to the research situation. The critical factor is how causally complex the phenomena being studied is. Woolcock (2022: 96-101) uses the term ‘causal density’, which refers to the degree to which the constituent elements of the process interact in similar ways across cases, or whether they are highly contingent based on the conditions present in a given case. We might know before engaging in PT that the phenomena is highly complex, or we might first identify this when engaging in comparative PT. Irrespective, if there is high causal density, we should in general be more cautious in making processual generalizations beyond the bounds of small subsets of cases. Further, the number of iterations of PT ‘light’ case studies in the extensive phase will be greater when exploring the bounds of processual homogeneity, other things equal.

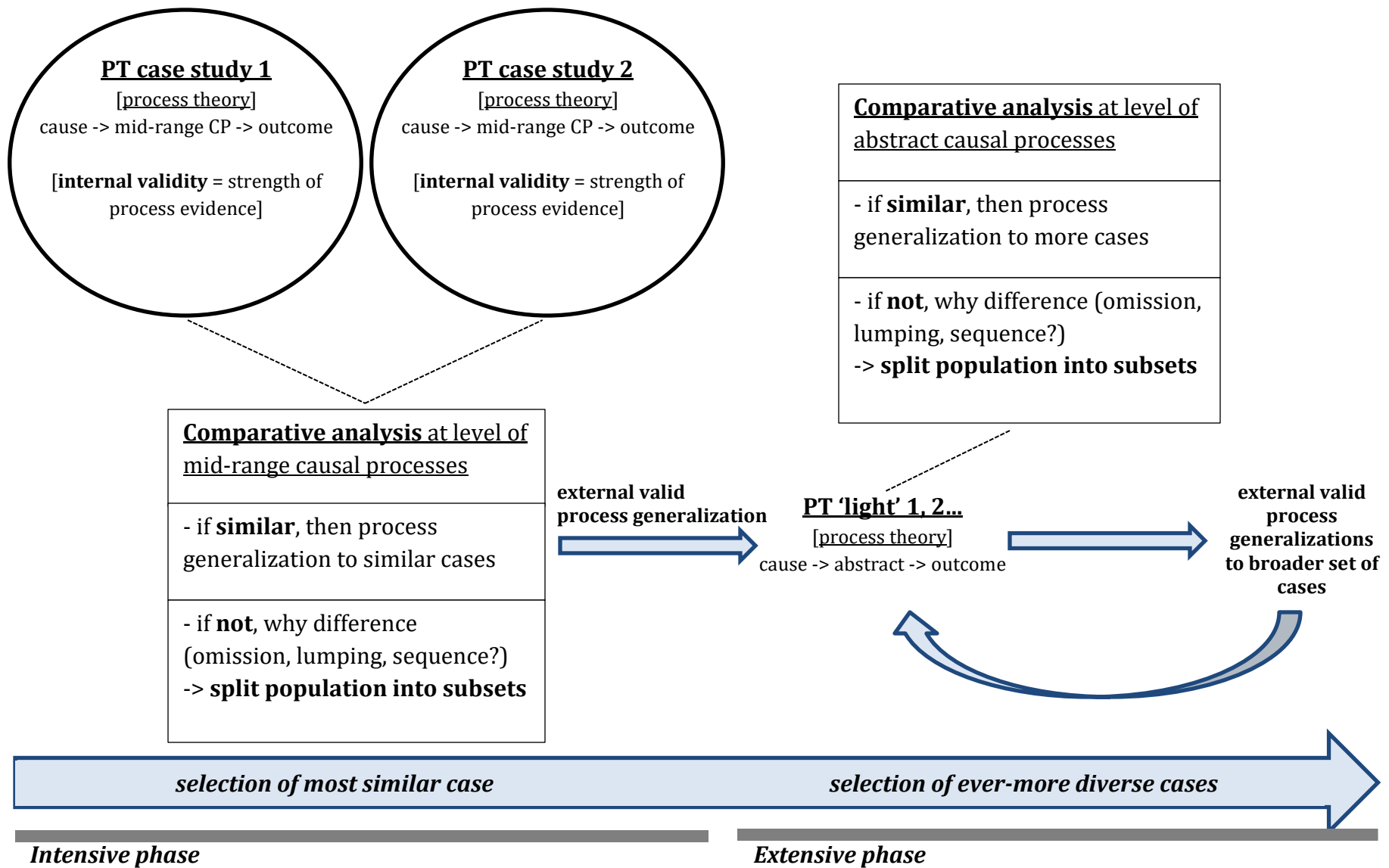


Figure 5 – Comparative Process Tracing.

Making things easier is the fact that we might not need to explore all of the potential reasons for processual heterogeneity in a study. For instance, there might be good theoretical (or empirical) reasons for focusing more on lumping than other sources of heterogeneity because we are concerned about how the cause was conceptualized. In another, we might be more concerned about known omission, and therefore focus on testing whether these omitted conditions/variables impact which processes are operative.

3.1. Intensive phase

The intensive phase of comparative PT involves the in-depth PT analysis of two or more cases. If similar processes are found in the two cases, the evidence-based generalization can be made that similar processes should be expected in other cases that are very similar. If different processes are found when comparing processes in the two cases, the comparison should turn to uncovering what is the difference that can explain differing processes. Is it due to omissions, lumping or sequences/interactions?

The first part of the intensive phase involves the initial PT case study. The first case should be selected based on several criteria. First and most importantly, the case should exhibit the outcome of interest, and if a known process is being traced, the cause(s) that might have triggered the process should also be present (see Beach and Pedersen, 2018). In other words, cases are selected based on Y being present, and often also X – with the goal of research to trace what chain of activities links them together (if any). In more practical, policy evaluation settings, the outcome to be selected might be policy ‘success’, with the subsequent evaluation attempting to figure out how the policy intervention was linked (if at all) with the outcome.

When we have information from cross-case analysis prior to engaging in the initial PT about case scores in relevant conditions/variables across a set of cases, we can use this information to select a case that is as similar as possible to the most other cases as a supplemental criteria. The logic here is that we should expect similar processes in similar cases, and when we are interested in processual generalizations, we should therefore start with selecting amongst cases that are similar to others. In table 3, we would want to select either A, B or D because there are at least two other cases that appear to be relatively similar on known

conditions/outcomes.¹⁴ Which of the cases A, B or D should we select? Here more practical considerations can come in, including access and linguistic issues.

	X1	X2	Y
A	1	0	1
B	1	0	1
C	1	1	1
D	1	0	1

Table 3 – Selecting ‘most-similar’ cases.

If we find evidence of a causal process in the first case, we would not be able to make any processual generalizations yet. However, by empirically tracing a mid-range process theory, we also shed light on the supporting and constraining conditions that *might* have to be present in order for the process to function as theorized (Bardach, 2004: 211; Cartwright and Hardie, 2012). Even a mid-range process theory should detail what activities are being performed by actors at critical stages of the process. The analysis thereby also sheds light on what conditions (might) have to be present in other cases when we ask ourselves what conditions have to be present for a given activity to be performed. For example, if an entrepreneur tabled a policy proposal, did they have to have a particular institutional role to be able to table it? Could another actor have done something similar without having the same role? Of course, here this is a form of counterfactual reasoning that is inherently speculative. At the same time, thinking about what supporting conditions might have to be present helps us move to a second case as we would want to select a case where the supporting conditions were also present. If the result of this theorization and reasoning suggests that there are very

¹⁴ - Note that we do not suggest that one selects cases where only one condition is present, as is suggested in the literature (Gerring, 2007; Schneider and Rohlfing, 2016). While isolating the effect of X1 and keeping everything else constant makes sense in experimental designs, if we are interested in processes, X2’s absence might impact what type of process is operative in case A. For instance, whether executive constraints are present or absent might change what types of processes are operative relating presidential impact on policy.

particular supporting conditions that cannot be expected to occur in other cases, the process might be very localized. However, given that this is *speculative*, we would still want to explore whether this actually holds empirically by exploring whether the process did work in other cases or not.

Using an example, the in-depth PT analysis by Winward (2021) found evidence for a simple, mid-range process theory involving collaboration with local elites that leads to an ever-worsening cycle of violence through four sets of activities. This is depicted in table 4, below.

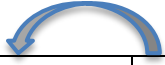
	Cause (trigger)	collaboration with local elites causal process				Outcome
						
causal process theory	Low intelligence capacity of security forces AND Civil conflict situation	Security forces approach local elites for information	Local civil elites provide false information targeted against specific group AND encourage civilian violence against members of group	Security forces torture leads to more arrests	Security forces mass detainment strains capacities, leading to increase in execution	Mass categorical violence
supporting conditions		- security forces lack other good sources - degree of trust between security forces and local elites	- local elites have incentives to exploit for reasons that do not overlap with causes of unrest - elites are relatively cohesive	- lack of rule-of-law constraints on security forces	- lack of rule-of-law constraints - poor prison conditions and prison capacity	

Table 4 – Mid-range causal process linking low intelligence capacity with violence.

Source: based on Winward, 2021.

Based on the article, we can reason that certain conditions might be required for the activities to function as theorized. For example, we might expect that underlying the provision of false information to security forces by elites would be the incentives that local elites in exploiting their role in supplying information. If they do not have strong pre-existing grievances, there would be no reason for them to supply false information. Additionally, if elites are not relatively cohesive, then we would expect that security forces might get very different information from different elite actors. Both can be expected to be important supporting conditions that would impact whether the collaboration with local elites functions as theorized in other cases also. Winward writes that, ‘These conditions ought to hold whenever security forces seek to deploy mass violence and require civilian collaboration to do so’ (Winward, 2021:578).

The second PT case should then be selected based upon both upon whether it is most-similar to the initial case (e.g. PT 1 selected case A, and PT2 then selected case B), and whether the speculative supporting conditions were present or not. Note that the selection of the second PT case typically involves figuring out whether the expected supporting conditions were present in other cases, which can require considerable case knowledge/research.

	X1	X2	supporting condition A	supporting condition B	Y
A [PT1]	1	0	1	1	1
B [PT2]	1	0	1	1	1
C	1	1	1	1	1
D	1	0	1	1	1
E	1	1	1	0	1
F	1	0	0	0	0

Table 5 – selecting a second case for in-depth PT.

The case selection strategy is depicted in table 5, in which case A was selected for the initial PT. Based on the process theory from case A, two potential supporting conditions (A and B) were expected to impact whether and how the process worked. The second PT case study could then select case B because it shares values on X1, X2 and A and B.

It could be argued that a most-diverse case should be selected for the second OT because it would cut right to assessment of the outer bounds of the operation of the process. However, this would be a high risk strategy for any but the simplest of causal processes. Given the many differences, we would expect different processes to be operative, and if we found evidence of another process, there would be so many differences that it would be very difficult to isolate the reason for the different processes. Instead, we suggest a more incremental, snowballing outwards strategy as a follow-up to the intensive comparative phase.

Comparing the two PT cases at the level of process enables evidence-based processual generalizations to be made, whereas if we only have evidence from one case, we have no way of knowing that similar process operative in any other case because of the three problems discussed in the previous section (lumping, omission, sequence and interaction). Of course, all but the simplest causal processes differ across cases at the level of a detailed, blow-by-blow account of the activities of actors. However, as discussed in section 3, when we move to the level of contingent or mid-range process theories, we can compare more abstractly defined activities during the critical phases of the process across cases.

If we find similar processes in case B, we can then compare the two cases in light of the three problems to assess whether there are any differences, based on the comparative logic that differences (at the level of conditions/variables) *cannot* explain similarities (at the process level). For instance, as regards assessing lumping, our comparison of the two similar cases might find that X1 was a family resemblance concept with two attributes, and that cases A and B have different attributes on this condition, thereby creating the risk of cases being lumped together conceptually that have different processes. If we found similar processes in

the two cases, we could then infer that (at least for these cases), there is not evidence that lumping was taking place as regards X1. There might also be known omissions. Our comparative analysis might have found that X3 did not matter in the cross-case analysis. If the two cases differ in values of X3, and the process is similar, there is evidence that X3 might not matter at the process-level also. Finally, there might be differences in sequences and interactions between similar cases. X1 might occur before ~X2 in case 1, and ~X2 before X1 in case 2. If we find similar processes, we have evidence that sequence might not matter at the level of process.

As an example, returning to the article by Winward (2021), he undertakes a second PT case study of a most-similar case (East Java). He finds confirming evidence that each critical phase of the process worked in roughly a similar manner to the first PT case. There were some differences, for instance the divisional commander in East Java was initially reluctant to coordinate an anticommunist campaign, meaning that any collaboration was at the discretion of local commanders. However, this changed after a brief period of time, and 'Following this, the dominant pattern of violence in East Java broadly resembled the joint operations described in Central Java' (Winward, 2021: 575). However, given that the two PT case studies only trace processes in very similar cases, it would be difficult to generalize much beyond the Java mid-1960s context. In order to make broader, evidence-based process generalizations, comparative PT would then move to the second, more extensive phase.

If we do not find evidence of a similar process in the second PT case,¹⁵ we would want to uncover the difference between the cases that might account for different processes. Here the comparative logic is that a difference can explain a difference. Cases might have been lumped together that are heterogeneous at the level of process due to how concepts are defined, or the cross-case analysis might have suggested that a condition/variable did not matter at the level of causes/outcomes, but it might impact how processes play out within cases. Finally, there might be different sequences or interactions between

¹⁵ - Note that if strong disconfirming evidence is found relatively early in the research for the second PT case study, it is not necessary to complete the PT analysis. Instead, the focus can shift towards identifying the difference between cases that can have produced different processes.

conditions/variables that impacts processes. Based on the cause of difference, the cases should then be re-coded into sub-sets within which different causal processes are operative.

If the second case does not find evidence of the process, after we have uncovered the difference and re-coded cases into subsets, we might then want to select another most-similar case for in-depth PT to assess whether the process is operative in other cases or not. If we again find that the process did not work in the same manner, there is increasing evidence that the process found in PT1 was idiosyncratic. However, even here we might want to explore whether it instead makes sense to lift the level of theoretical abstraction a bit higher when comparing the processes. If they still differ, then the process in PT1 might indeed be particular to that case.

3.2. The extensive phase

The extensive phase uses the comparative analysis of ever-more diverse cases to explore empirically the bounds of valid processual generalizations, asking in effect how far can it travel. Note that it is not a 'mechanical' approach, but instead choices should be based on theoretical/empirical knowledge of the phenomena being studied when deciding which potential differences might produce different processes (lumping, omissions, sequences and interactions). For instance, we might be particularly concerned that parliamentary dynamics might differ depending on electoral rules despite our cross-case analysis telling us that different rules did not produce different outcomes. PT case studies could then explore whether similar processes were operative in countries with more/less proportional representative systems.

For practical reasons, we suggest that PT in the extensive phase shifts towards what can be thought of as a form PT 'light', in which the focus is on assessing whether one or more key empirical signatures from one or more critical phases is present or not in any given case (Steel, 2008: 88-92; Beach, Pedersen, and Siewert 2019:133-45). Note that we are not advocating the use of mediation analysis because this moves to the realm of large-n cross-case analysis (Imai et al, 2010), in which we move so far away from individual cases that it

becomes difficult to return to the comparative analysis of two cases to explore why there are different processes at play.

Cases can then be selected that are ever-more diverse cases can be used to explore whether similar processes are operative or not. If we find similar processes, then the difference between the already studied and additional case can be 'eliminated', enabling process generalizations to cases where the eliminated condition is either present or absent. If different processes are found, comparisons are undertaken to detect why there was a difference, and the newly found difference is used to bound cases into processual homogeneous subsets.

How the extensive analysis actually proceeds depends on the type of difference being probed, which is a function of the causal complexity of the phenomena being analyzed (Woolcock, 2022). Potential lumping can be detected by assessing cases on conditions that might hypothetically make a difference for process. For instance, we might be concerned that a differences-in-degree on X might trigger different processes (high values might trigger a different process than intermediate values). Known omissions can be probed by selecting cases that differ on conditions/variables dropped by the cross-case analysis. Given that we do not know what we do not know until we find it, if we do not find different processes across increasingly diverse cases, the risk of unknown omissions is gradually reduced, but never eliminated. If we find different processes in cases that otherwise look similar, we have detected an unknown omission that should be identified through careful comparison of two or more cases. Sequences and interactions can be probed through the selection of cases where there are differences in the sequencing of conditions/variables, and/or where there are plausible interactions between conditions/variables (e.g. several conditions are found by QCA to be sufficient for the outcome).

Table 6 illustrates the selection of ever-more diverse cases. After the intensive phase, we might be concerned about a known omitted condition (X3), and therefore select case C to do a PT 'light' to explore if there is any evidence suggesting that X3 makes a processual difference. If we processual differences, we could then delimit the population to which the

process works to cases similar to A and B. If we find no processual difference, then case D could be selected to probe whether the supporting conditions really have to be present in all cases or not. If we find a similar process here, we can infer that supporting condition B is not always necessary, and additionally we might search for additional (unknown) supporting conditions that might be present across all of the cases A to D.

	X1	X2	X3	supporting condition A	supporting condition B	Y
<i>intensive phase</i>						
A [PT1]	1	0	0	1	1	1
B [PT2]	1	0	0	1	1	1
<i>extensive phase</i>						
C	1	0	1	1	1	1
D	1	0	1	1	0	1
...	<i>ever-more diverse cases...</i>					
<i>deviant consistency case</i>						
E	1	1	1	1	0	0

Table 6 – Case selection in the different phases of comparative PT.

For instance, returning again to the Winward (2021) example, there are two cases that we might select as more diverse cases: Columbia from 2002-8, and Guatemala from 1960-96. In Columbia, according to the most recent estimates, 450,664 people lost their lives during the full period of armed conflict (1958 – 2019), of which 45% happened between 1995 and 2004 (Colombia Comisión de la Verdad, 2022:127). In the period between 2002 and 2008 there was relatively low intelligence capacity (X) and the outcome of categorical violence was also

present in the deliberate targeting of innocent people (so-called 'falsos positivos'), with 6,402 documented victims of extrajudicial executions during the time period (Colombia Comisión de la Verdad, 2022:131), making this time period a relevant case to engage in PT. There were some obvious differences between the cases. In contrast to the Indonesian cases, the conflict in Columbia was a civil war, but with the similarity that 90% of the victims (including enforced disappearances) were civilians, and that paramilitary groups played an important role in the violence against civilians (ibid:127). Another obvious causal difference between Indonesia and Columbia would be the role that drug trafficking played on both sides of the conflict.

Another potential case could be Guatemala during the time period of 1960 to 1996, in which both low intelligence capacity and high levels of categorical mass violence were present. Guatemala's truth commission estimated the number of assassinations to be around 132,000, and they believe that multiple genocides took place (Guatemala Comisión para el Esclarecimiento Histórico, 1999a:71, 199b:314). As with the Columbia case, the categorical violence was embedded in a civil war in contrast to more localized unrest in Indonesia. Another difference was the role of the United States as an external third-party in the conflict.

If we then undertook a PT 'light' analysis of the Colombian case, we would quickly find that that the dynamics were quite different, especially because there not extensive collaboration with local elites, meaning that this part of the process linking X and Y was different, although the later escalatory dynamics with increasing violence due to torture etc were more similar.

In the 'falsos positivos' case in Colombia, the Colombian army relatively indiscriminately targeted vulnerable groups instead of the FARC insurgents or drug traffickers because they were rewarded for fighting irrespective of whether the people they apprehended (and usually executed) were actually insurgents or traffickers. This could even involve misleading people by offering poor people employment opportunities and then luring them to another location, at which they were killed and passed off as insurgents or traffickers. That these executions were considered a sign of success for the armed forces and the lack of investigations and consequences for the perpetrators created a different set of incentives for

army members. Instead of collaboration with local elites, the increase in violence was being driven by the incentives given to the army. That incentives were driving the process can also be seen in the fact that after the practice became publicized, 17 high ranking officials were removed from the army and the practice was stopped, resulting in a marked of 'falsos positivos' the following year (Colombia Comisión de la Verdad, 2022:132). These incentives could be a known omission, because the institutional incentives of the security forces were not deemed relevant in the Indonesia cases, but might matter in Columbia (and other cases). Further, there might be other differences such as how the history of the conflict impacted how actors behaved during the case. The 'falsos positivos' case might be better understood as a specific episode of the conflict in which the iteration and repetition of the violence produced particular behaviors. For example, because it was an ongoing civil war in which indiscriminate violence had been used repeatedly, this repeat nature made it easier for them to resort to violence again in the case. This repeat nature can have produced different causal dynamics in the process, making it a potential omitted condition that could be relevant in understanding the types of cases in which collaboration with elites occurs.

Similarly, a PT 'light' analysis of the Guatemala case would find that there was a different process operative. In Guatemala, security forces were so afraid of civilians who had loyalties towards insurgents that they targeted civilians as a way to deter them from collaborating with the insurgency in the future (Schwartz and Straus, 2018: 223). In Guatemala, the causal process could have been different due to the characteristics of the insurgency as well as due to historical factors in the region. For instance, after the Cuban Revolution, Central American saw an increased intervention of the United States on its local affairs. Although the conflict started in the 1960s, the largest escalation of violence by the security forces occurred in the 1980s. This appears to coincide with an increased effort by the insurgency of recruiting in the Mayan highlands; a territory already perceived by the government as an obstacle to national integrity (Schwartz and Straus, 2018:228). These indigenous populations were not considered part of the core national identity by the state, which allowed the violence against them to escalate (ibid:228). In this case, we see that the type of insurgency and its specific actions are more diverse than what we would have thought before analyzing the causal process in the case. The type of insurgency and its chosen strategy created a different

interaction with the security forces, leading to a different causal process. Whereas the characteristics of the PKI in Indonesia could be omitted as a relevant contextual condition, they impacted how the process worked in the Guatemala case.

The comparative processual analysis might then look like the following (table 7).

Cases	Cause	Part 1	Part 2	Part 3	Part 4	Outcome	Difference across cases
Central Java (Winward, 2021)	Low intelligence capacity	Security forces turn to elites for information	Elites widen targeting criteria	Torture and more detainees	Security forces use lethal violence to reduce number of people in prison	Mass categorical violence	
Colombia	Low intelligence capacity	Security forces deployed to regions far from the center	Government issues a reward system for combatting insurgents	Security forces and paramilitary groups use indiscriminate violence in order to show 'results' to their commanders in the city	Innocent people are targeted and killed or disappeared	Mass categorical violence (?)	-incentives of armed forces -repeat nature of conflict (civil war)
Guatemala	Low intelligence capacity	Security forces consider all civilians in the region to be loyal to the insurgent movement	Security forces target civilians	Violence used as a means to deter civilians from collaborating with the insurgence		Mass categorical violence	-nature of the insurgent recruitment -role of US as external third party

Table 7 – Comparative processual analysis of ever-more diverse cases.

Note – parts shaded grey differ across cases.

Both the Colombian and Guatemalan cases therefore have different causal processes linking the cause – low intelligence capacity – and the outcome – mass categorical violence. In the context of internal conflict and social unrest and low intelligence capacity, and a resulting violent cycle, we see that the armed forces behaved differently despite similarities in

cause/outcome. In neither case did the armed forces appear to have turned to the local elites to obtain information regarding the members of the targeted group and the escalation of indiscriminate violence against innocent people seems to result from a different causal process. In sum, although low intelligence capacity in the context of social unrest or an armed conflict in countries with weak rule of law appears to lead to innocents being targeted by the security forces, the collaboration with elites causal process appears to depend of more factors. It would therefore not have been valid for Winward (2021) to generalize that the collaboration with elites causal process worked in other cases outside of the context of Java in the 1960s. Instead, it appears there are different subsets of cases in which different supporting conditions are at play that lead to different causal processes. Further process-focused research on this topic could then explore whether there are any cases that are similar to the Java cases as regards the collaboration with elites process, or whether the process is so contextualized that it only worked in that way there. Alternatively, research could shift towards whether there are similar processes to Columbia or Guatemala present in other cases that share structural similarities with them, and if so, what are the bounds within which those processes are operative.

Once the comparative PT has explored the most plausible potential reasons for processual heterogeneity, deviant consistency cases can also be relevant to explore for omitted conditions/variables.

3.3. Analyzing deviant consistency cases

After we have developed a robust understanding of the processes at play across a series of positive cases, a last step could be to investigate a deviant consistency case, i.e. one where Y is not observed despite the fact that the main explanatory condition(s)/variables are present, and the known supporting conditions are present. By tracing a process until its break down, one can obtain important information about conditions which have to be present in order for a given process to work (e.g. Beach and Pedersen 2019:274–75; Goertz 2017:66–68)

The analysis proceeds in two steps. First, a process that we have evidence of in other cases is traced in the deviant case until the point at which it breaks down (Anderson 2011: 421–22). For instance, we might find out that a negotiation breaks down at the point at which the appointed chair tabled a compromise proposal. Second, the analysis then turns to a comparison of the breakdown case with a similar case where the process worked, attempting to assess what potential difference (or differences) could account for the difference in process (breakdown/worked). In this example, we might find that other parties did not think the chair was impartial, and therefore they rejected the proposal. Perceptions of impartiality might then be an unknown omitted condition that would differentiate cases where the process worked from breakdown.

A good example on how a deviant case can be used is found in Löblová's PT case study of the impact of epistemic (expert) communities on government's policy outputs (2018). In her analysis she studies the influence of experts in two cases: the typical case of health technology assessment in Poland, and the deviant constituency case of the Czech Republic for the same policy. In the Czech case, she finds that the process of expert influence breaks down because of changing conditions that is identified as a drop in policy-maker demand for expertise in the case.

Section 5 – Conclusions

Comparative PT as developed in this article is a methodological tool to enable evidence-based generalizations to be made about how causal processes operate in different cases. At its core comparative PT involves the process-level comparison of cases, and using information about processual differences or similarities to empirically explore the bounds of cases to which valid process generalizations can be made. Existing PT studies either engage in one-off analysis, which per definition is not cumulative knowledge, or make one-to-many generalizations that thereby reduce the analytical utility of follow-up analysis in other cases. In contrast, comparative PT has the ambition of enabling cumulative process-focused research by probing empirically through multiple case studies when similar or different processes are operative in different types of cases. Realistically, a given article need not fully engage in all of the steps of comparative PT, but instead can be embedded in a broader research program aimed at assessing under what conditions given causal processes can be expected to be operative. Ragin captured the ambition of empirically validating the bounds of valid generalizations when he wrote that, ‘...in diversity-oriented research, by contrast, population boundaries are not taken for granted, nor are they fixed. Instead, they are fluid. They can be revised up until the very end of a research project, as the investigator’s knowledge of cases grows and deepens.’ (2000: 37-38).

The article unpacked three different reasons why it is problematic to assume causal homogeneity at the level of causal processes across cases, including known/unknown omissions, lumping, and sequences and interactions. We then differentiated the phases of comparative PT into an intensive and extensive phase. The intensive phase allowed us to understand the conditions that triggered a particular causal process in our case. In this more extensive phase, conducting PT ‘light’ of two cases uncovered known and unknown omissions which now allows us to increase the external validity of our theory.

We recognize that comparative PT requires sustained attention and significant analytical resources. However, just because it is difficult does not mean that we should not do it, as the alternative is either to engage in one-offs, or merely assume that similar processes are

present without any empirical justification (one-to-many generalizations). While it is of course important for academics to understand how things work, making evidence-based generalizations is arguably much more important when thinking about the implications that research can have in terms of real-world policy-making. Policy-makers want to have 'evidence-based' knowledge when making new policies, but if the 'evidence' that we give them based on in-depth case studies is only internally valid, there is the risk that they apply the lessons learned to cases in which different causal processes might be operative. We should therefore strive to make clear when we have evidence about the types of cases within which a particular causal process can be expected to be operative, and the strength of the evidence (both in terms of internal and external validity). In this respect, it is important to be clear both about *what we know*, but equally important, signal clearly *what we do not know* both in terms of how things work and evidence suggesting where it might work.

5. References

- Bardach, E. 2004. Presidential address – The extrapolation problem: How can we learn from the experience of others? *Journal of Policy Analysis and Management*, 23(2): 205-220.
- Barrenechea, R., and Castillo, I. 2019. 'The Many Roads to Rome: Family Resemblance Concepts in the Social Sciences.' *Quality and Quantity* 53 (1): 107–30.
- Baumgartner, M., and Epple, R. 2014. A Coincidence Analysis of a Causal Chain: The Swiss Minaret Vote. *Sociological Methods and Research*, 43(2): 280–312.
- Beach, D. 2020. 'Multi-Method Research in the Social Sciences: A Review of Recent Frameworks and a Way Forward.' *Government and Opposition* 55 (1): 163–82.
- Beach, D. 2022. 'Process tracing methods in the social sciences.', *Oxford Research Encyclopedia of Politics*, <https://oxfordre.com/politics/view/10.1093/acrefore/9780190228637.001.0001/acrefore-9780190228637-e-176>
- Beach, D., and J. G. Kaas. 2020. 'The Great Divides: Incommensurability, the Impossibility of Mixed-Methodology, and What to Do about It.' *International Studies Review* 22 (2): 214–35.
- Beach, D., and R. B. Pedersen. 2016. *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching, and Tracing*. Ann Arbor, MI: University of Michigan Press. <http://www.press.umich.edu/6576809>.
- Beach, D. and R. B. Pedersen. 2019. *Process-Tracing Methods: Foundations and Guidelines*. Second Edition. Ann Arbor: University of Michigan Press.
- Beach, D., and Rohlfing, I. 2018. 'Integrating Cross-Case Analyses and Process Tracing in Set-Theoretic Research: Strategies and Parameters of Debate.' *Sociological Methods and Research*, 47 (1): 3–36.
- Bechtel, W. and Richardson, R.C. 2010. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. MIT Press.

Bennett, A., and Checkel, J. (Eds.). 2014. *Process Tracing: From Metaphor to Analytic Tool* (Strategies for Social Inquiry). Cambridge: Cambridge University Press.

Binder, M. 2015. 'Paths to Intervention: What Explains the Un's Selective Response to Humanitarian Crises?' *Journal of Peace Research* 52 (6): 712–26.

Bretthauer, J. M. 2015. Conditions for Peace and Conflict: Applying a Fuzzy-Set Qualitative Comparative Analysis to Cases of Resource Scarcity. *The Journal of Conflict Resolution*, 59(4): 593–616.

Cartwright, N. 2011. 'Predicting 'It Will Work for Us': (Way) Beyond Statistics.' In *Causality in the Sciences*, edited by McKay Illari, P.; Russo, F., and Williamson, J. 750–68. New York: Oxford University Press.

Cartwright, N. 2020. Middle-range theory: Without it what could anyone do? *THEORIA. An International Journal for Theory, History and Foundations of Science*, 35(3): 269-323.

Clarke, B.; Gillies, D., Illari, P.M., Russo, F., and Williamson, J.. 2014. 'Mechanisms and the Evidence Hierarchy.' *Topoi* 33 (2): 339–60.

Collier, D., and Levitsky, S. 1997. 'Democracy with Adjectives: Conceptual Innovation in Comparative Research.' *World Politics* 49 (3): 430–51.

Colombia, Comisión de la Verdad. 2022. Hay futuro, si hay verdad: Informe Final de la Comisión para el Esclarecimiento de la Verdad, la Convivencia y la No Repetición. Vol. II. Hallazgos y recomendaciones. Bogotá: Comisión de la Verdad.

Craver, C.F., and Darden, L. 2013. *In Search of Mechanisms*. Chicago: The University of Chicago Press.

Craver, C.F., and Kaplan, D.M. 2020. Are More Details Better? On the Norms of Completeness for Mechanistic Explanations. *The British Journal for the Philosophy of Science*, 71(1):287-319.

- Elster, Jon. 1998. A plea for mechanisms. In *Social Mechanisms*, ed. P. Hedström and R. Swedberg, 45-73. Cambridge: Cambridge University Press.
- Fairfield, T. and Charman, A.E. 2017. Explicit Bayesian Analysis for Process Tracing: Guidelines, Opportunities, and Caveats. *Political Analysis*, 25(3): 363-380.
- Falleti, T. G., and Lynch, J. F. 2009. 'Context and Causal Mechanisms in Political Analysis.' *Comparative Political Studies* 42 (9): 1143-66.
- Falleti, T. G., and Mahoney, J. 2015. 'The Comparative Sequential Method.' In *Advances in Comparative-Historical Analysis*, edited by James Mahoney and Kathleen Thelen, 211-39. Cambridge; New York: Cambridge University Press.
- Gerring, J. 2010. 'Causal Mechanisms: Yes, But....' *Comparative Political Studies* 43 (11): 1499-1526.
- Gerring, J. 2007. 'Is There a (Viable) Crucial-Case Method?', *Comparative Political Studies*, 40(3): 231-253.
- Glennan, S., and Illari, P. Eds. 2017. *The Routledge Handbook of Mechanisms and Mechanical Philosophy* (1st ed.). Routledge.
- Goertz, G. 2006. *Social Science Concepts*. Princeton: Princeton University Press.
- Guatemala Comisión para el Esclarecimiento Histórico. 1999. Guatemala, memoria del silencio. Guatemala: UNOPS.
- Haggard, S. and Kaufman, R.R. 2016. *Dictators and Democratic: Masses, Elites and Regime Change*. Princeton University Press.
- Imai, K., L. Keele, D. Tingley, T. Yamamoto. 2011. 'Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies.', *American Political Science Review*, 105(4): 765-789.

- Jackson, P. 2016. *The Conduct of Inquiry in International Relations*. 2nd Edition. London: Routledge.
- Khosrowi, D. 2019. 'Extrapolation of Causal Effects – Hopes, Assumptions, and the Extrapolator's Circle.' *Journal of Economic Methodology* 26 (1): 45–58.
- Kuehn, D., and Rohlfing, I. 2009. 'Does It, Really? Measurement Error and Omitted Variables in Multi-Method Research.' *Qualitative and Multi-Method Research* 7 (2): 18–22.
- Kuehn, D.; and Trinkunas, H. 2017. Conditions of military contestation in populist America. *Democratization*, 24(5):859-880.
- Löblová, O. 2018. 'When Epistemic Communities Fail: Exploring the Mechanism of Policy Influence: When Epistemic Communities Fail.' *Policy Studies Journal* 46 (1): 160–89.
- Mahoney, J. 2015. Process Tracing and Historical Explanation. *Security Studies*, 24(2): 200-2018.
- Mello, P. A. 2012. Parliamentary Peace or Partisan Politics? Democracies' Participation in the Iraq War, *Journal of International Relations and Development* 15(3): 420-53.
- Møller, J., and Skaaning, S.E. 2010. 'Beyond the Radial Delusion: Conceptualizing and Measuring Democracy and Non-Democracy.' *International Political Science Review* 31 (3): 261–83.
- O'Mahoney. 2017. Making the Real: Rhetorical Adduction and the Bangladesh Liberation War. *International Organization*, 71(2):317-348.
- Ragin, C. C. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: University of Chicago Press.
- Ragin, C. C., and Strand, S. I. 2008. Using Qualitative Comparative Analysis to Study Causal Order: Comment on Caren and Panofsky (2005). *Sociological Methods and Research*, 36(4): 431–441.

Ross, M. L. 2004. 'How Do Natural Resources Influence Civil War? Evidence from Thirteen Cases.' *International Organization*, 58(1), 35–67.

Salmon, Wesley. 1998. *Causality and Explanation*. Oxford: Oxford University Press.

Sayer, Andrew. 2000. *Realism and Social Science*. London: Sage Publications.

Schneider, C. Q., and Rohlfing, I. 2016. 'Case Studies Nested in Fuzzy-Set QCA on Sufficiency: Formalizing Case Selection and Causal Inference.' *Sociological Methods and Research* 45 (3): 526–68.

Schwartz, Rachel A., and Scott Straus. 2018. 'What drives violence against civilians in civil war? Evidence from Guatemala's conflict archives', *Journal of Peace Research*, 55(2): 222-35.

Steel, D. 2008. *Across the Boundaries: Extrapolation in Biology and Social Science*. Environmental Ethics and Science Policy Series. Oxford ; New York: Oxford University Press.

Waldner, D. 2014. 'What makes process tracing good?' In *Process Tracing: From Metaphor to Analytic Tool* (Strategies for Social Inquiry) edited by Bennet, A. and Checkel J. Cambridge: Cambridge University Press.

Waldner, D. 2015. 'Process Tracing and Qualitative Causal Inference.', *Security Studies*, 24(2): 239-250.

Wilde, M., and Parkkinen, V.P. 2019. 'Extrapolation and the Russo–Williamson Thesis.' *Synthese* 196 (8): 3251–62.

Winward, M. 2021. Intelligence Capacity and Mass Violence: Evidence From Indonesia. *Comparative Political Studies*, 54(3–4): 553–584.

Woolcock, M. 2022. 'Will It Work Here? Using Case Studies to Generate 'Key Facts' About Complex Development Programs.', in Widner, Woolcock and Nieto (eds) *The Case for Case Studies: Methods and Applications in International Development*. Cambridge: Cambridge University Press, pp. 87-116.

Xie, Y., Brand J.E., and Jann, B. 2012. 'Estimating Heterogeneous Treatment Effects with Observational Data.' *Sociological Methodology* 42 (1): 314–47.