Learning from Cases – comparative process tracing and the study of crisis in the EU

Authors:

Derek Beach, Aarhus University (derek@ps.au.dk)
Sandrino Smeets, Radboud University (sandrino.smeets@ru.nl)

DRAFT – DO NOT CITE

Paper prepared for the biannual meeting of the European Union Studies Association, Pittsburgh, PA, USA. May 4-6, 2023

1. Introduction

Scholars of European integration and governance who engage in case study research confront the challenge of how one can learn anything 'general' about how things work from in-depth research on one or a small handful of cases.[1] All too frequently scholars conclude an in-depth case study with a discussion of the implications of the findings for our 'general' understanding of a given phenomena, but where the one-to-many generalization is not supported by any actual evidence that the process works in similar ways in other cases. This is problematic because no two cases are *identical* when we are studying complex phenomena such as how the EU manages crisis (e.g. euro-crisis, migration crisis, COVID-19, etc), the processes leading to major treaty reforms being adopted (e.g. Single European Act, Maastricht, Amsterdam, etc), or how policy processes are managed that produce major legislative packages (e.g. Single European Market, European Green Deal, etc.). Learning about how things work at the process level (aka within-case analysis) based on in-depth case study research therefore creates two challenges.

First, how can we establish that similar processes were at play across two or more cases? Given that every case is unique at the empirical level, how can we ever establish similarity in terms of how things worked *across* cases? While it is relatively easy to establish that input (causes, e.g. crisis) and output (outcomes, e.g. reform) are similar in two or more cases through various forms of comparative analysis, how can we ever establish that causal process(es) worked in similar ways? For example, at the empirical level, the 1990-91 IGC's that led to the Treaty of Maastricht was very different from the previous round of major treaty reform in the 1985 IGC. How could we ever establish that institutional actors played similar instrumental leadership roles in the two rounds of reform? How can we assess processual similarity, and on what parameters should the assessment of similarity be based on? While there exists a large body of case study methodological literature that deals with

[1] - A 'case' is defined in this article as one instance of a causal theory playing out, in terms of Cause (X) -> causal process -> Outcome (Y). When selecting cases for tracing how causal processes worked, we obviously only have knowledge of a cases scores on the cause(s), outcome, and sometimes some knowledge of relevant scope conditions that have to be present for a given process to work.

how to make within-case inferences (aka internal validity), there exists little guidance for how to engage in comparative analysis at the process level (aka external validity).[2]

Second, if we evidence processual similarity across two cases, how can we know that similar things are at work in other, unstudied cases?[3] How can scholars avoid the extremes of either: a) making 'one-size-fits-all' type theories that can be expected with high confidence to be applicable in all other potential cases, but where the theory tell us next to nothing about how things actually worked in any given case?, or b) settling for making no 'generalizing' inferences beyond the studied case(s), for example by studying *all* potential cases? Note that even if one does not have the ambition to 'generalize', as soon as a theoretical abstraction is made based on findings from a case, the theory becomes at least an implicit generalization that might be applicable to other, unidentified cases.

This article makes a practical contribution to the literature on European integration by offering case study scholars methodological guidelines for making evidence-based generalizations about how things work for a given phenomena within cases. The first part of the guidelines deals with how can we establish that similar processes are operative across two cases. This involves selecting a first case in which 'abstracting' is undertaken, in which the essence of the causal process is distilled from the empirical record in terms of *who* (key actors), *what* and *how* (key activities and causal principles linking them together) during *key episodes*. Note that a good process theory includes enough processual attributes (who, what, when and how) that it is possible to establish whether the process worked in similar or different ways in other cases.

After a good process theory is developed, another case that is (relatively) similar to the initial case is then selected based on scores on the cause(s), outcome and plausible contextual (aka scope) conditions. The process theory developed in the first case is then 'particularized' in relation to the empirical context of the new case, with the scholar deploying deep case

[2] - One of the authors has worked on issues of internal validity of process tracing methods for over a decade, but recognizes that the issue of external validity has for the most been overlooked.
[3] - Unstudied cases can be both those that have already occurred, but that might also occur in the near future.

knowledge of both cases to assess whether there was functional equivalence at the process level in terms of who, what and how in key episodes linking the cause and outcome together. For example, in the case of managing the banking crisis in 2012-14, the German government played the role of a laggard, veto-player, whereas in the migration crisis in 2015-16, the French government might have played a similar (but not identical) role in the reform process. Establishing this type of functional equivalence in processes requires considerable empirical case knowledge for both cases.

The second part of the framework deals with exploring 'to what' the processual theory might be applicable. Often scholars engage in the study of one or two cases and then discuss the 'broader implications' for European integration, implying that the theorized process drawn can be expected to be present in all other cases. However, this would be a generalization based on hope, in which the scholars *assumes* that findings of how it worked 'here' means that it should work that way 'everywhere' without any actual empirical evidence. We put forward a snowballing outwards procedure that enables scholars to explore whether similar (but never identical!) processes are at play in ever more diverse cases, thereby enabling stronger *evidence-based* generalizations to be made while signaling the level of confidence we can have in them.

In this article, we utilize our own research program that has been published in a series of articles to illustrate the framework. We have been interested in tracing the processes triggered by major EU crises, in which the European Council informally delegates responsibility for developing and negotiating reforms to a network of institutional actors, with the result (in some cases) that final reforms are agreed more quickly, and are more ambitious than member states left to themselves would have achieved. However, given the difficulty of cramming even a small proportion of the evidence backing our processual claims into the space constraints of a journal article, each article has had a stand-alone character in which we present an abstract process theory and the evidence backing it. But then each article begs the question of whether the knowledge produced about how things worked in the given crisis case tells us anything more 'general' about other crisis cases. Can we take what we learned from how informal delegation to institutions processes worked in the case

of the Fiscal Compact and use it to shed light on other crises, e.g. the Covid-19 response? If so, how do we then avoid engaging in purely theoretical speculation along the lines of I have evidence of how it worked here, ergo it should work like that everywhere without any empirical evidence suggesting that it works in similar ways in other cases? In this article, we show the types of generalizing claims we have made, and illustrates how the clarity of the evidential basis for these claims could be improved by using the framework more explicitly in cumulative, process-focused research of multiple cases.

2. The challenges of making processual generalizations about how things work

Theories are *per definition* abstractions from particular empirical cases,[4] in which formal nouns and case-specific verbs are replaced with more abstract concepts and causal linkages that can in principle be present in other cases (Przeworski and Teune, 1966). Theories are therefore always a form of generalization. We theorize to be able to make sense of the complexity of a particular case, but the act of abstracting from the particular then opens the possibility that the theory might also be at play in other, similar cases in a population.[5] A population of cases is defined in this article as the domain (or scope) within which a given cause->process->outcome works, with a good theory provides us with enough detail that it enables us to distinguish between cases where it works and where it did not.

Whether the scholar makes it explicit or not, generalizations always refer to a *target population* within which the theory might in principle be operative. In other words, a theory always includes a 'to what target population' element. Unfortunately, many studies move back-and-forth between the particular and general (abstract theory) with little consideration for what is being compared (if at all), and the warrant for making 'general' claims.

This can take many forms. For instance, in a theory section, scholars are do not explicitly identify the relevant target population within which they are theorizing (e.g. by not discussing explicit scope conditions), and/or by not referring explicitly to the empirical

[4] - Note that the core challenge of moving from the particular to the abstract, and back again, is a common challenge shared by ALL methods. A large-n survey is an empirical observation of a sample of a population. While the sample is hopefully representative of the population, it is not representative across space and time, meaning that a single survey is no more externally valid than a single case study.

[5] - The term 'population' is ambiguously used in the broader methodological literature. In large-n, quantitative research, it is often used to refer to statistical sampling (sample in relation to a population). However, here this is only the potential population at a specific point in time and space. Yet the population to which a 'general' causal inference refers is typically much larger in space and time. A study that uses a survey of a statistically representative sample of the EU voting age population at a given point in time still lacks external validity in relation to the potential population of individuals across space and time (e.g. voters in the EU-15 in the 1990s versus voters in the current EU-27). Population is therefore used in this article to refer to the universe of potential cases to which a given causal claim can be operative.

evidence from studies of particular cases that (hopefully) exists that demonstrates that a theory works within a given population of cases. Particular case studies often do not situate the selected case(s) within a target population when discussing case selection. Sometimes scholars put forward such simple, 'one-size-fits-all' type theoretical claims about how things work (e.g. consensus-seeking dominates EU decision-making) that they are in principle generalizable to a large set of cases (e.g. all legislative processes since the EU's founding?), but where the claim tells us next to nothing about any individual legislative process. Does consensus-seeking include all actors, or only governments in the Council of Ministers? How does consensus actually work in producing outcomes? Who is doing what, when and why? While putting forward such a broad claim is not without merit, it leaves us in the dark as regards what we should be tracing empirically in any given case, and it offers little in terms of being able to assess similarity/differences between how things work in different cases. The need for including more attributes in process theories to enable differentiation between cases echoes Sartori's call for including more attributes when defining concepts to avoid 'no difference' universals that do not enable us to detect similarities and differences between cases (1970). Finally, when discussing the 'implications' of the findings many articles engage in relatively loose speculation that similar things should be at work in other cases without providing any actual evidence backing the generalization.

There are (luckily) numerous examples of scholars engaging in current methodological best practices by either: 1) not engaging in pure speculation that what was found in case(s) is 'general', resulting in a 'one-off' type study, or 2) engaging in some form of comparative processual analysis, but where a strong methodological frameworks that enables adjudication of process similarity/difference is lacking, as is the basis for making evidence-based generalizations. For instance, Jones et al (2016) conclude their case study of 'failing forward' in relation to the eurocrisis with the following:

> 'We expect that "failing forward" can be used to explain political developments in many other areas of European integration, such as immigration policy, and we hope that future studies will explore whether and to what extent such dynamics are at work in other fields.' (p. 1027)

Here they suggest that the 'failing forward' process evidenced in the case study might be present in other cases, but they follow current best practice by not making a strong generalizing claim. In a later special issue of *Journal of European Public Policy* in 2021 (vol. 28, issue 10), a range of studies by different authors are deployed to attempt to provide evidence behind the claim that 'such dynamics are at work in other fields'. However, these comparisons lacked a clear set of parameters upon which similarity could be assessed. For example, they admit in the introduction of the special issue that concepts in their theory such as 'failure' and 'forward' are ambiguous. This makes it difficult to assess whether the trigger of the rest of the process is actually present or not in any given case. Most problematic, they do not unpack the process theory beyond a simple abstract pattern, [6] making it very difficult for the studies to assess whether similar processes were actually operative across cases.

With current methodological best practice, it is difficult to assess whether similar processes were actually present in different cases because there is a lack of guidance for what parameters to assess similarity on. When are two or more cases 'similar enough'? And if we find similar processes in two cases, when and how can we say something more 'general' about other, unstudied cases?

One solution is to avoid the problem by not making any generalizing claims on the basis of a case study. However, this would require that the case study is a thick descriptive account without any form of explicit theorization.[7] As discussed above, theorization is a generalization away from the particular, even if the population to which the generalization refers might not be made explicit.

Another solution is to engage in more strategic case selection, but this does not solve the problem of a lack of evidence-based generalizations, nor does it provide a vocabulary for

[6] - '...in an initial phase, lowest common denominator intergovernmental bargains led to the creation of incomplete institutions, which in turn sowed the seeds of future crises, which then propelled deeper integration through reformed but still incomplete institutions – thus setting the stage for the process to move integration forward...' (Jones et al, 2021: 1519-20).

[7] - This would mean that the case study is an idiographic, historical account of what happened in a given space and time. We do not claim that this type of work has no utility. On the contrary, detailed empirical accounts are vital for understanding important historical episodes in the EU.

drawing comparative processual lessons from cases. Strategic case selection involves a comparative mapping of a population of potential cases on values of causes, outcomes and scope/contextual conditions. One strategy involves distribution-based selection of cases. For instance, a diverse case selection (Gerring and Seawright, 2007) would involve selecting two cases on scores of causes and outcomes (e.g. a high/high and low/low), and then based on finding processual similarity(?) making the assumption that it works that way in other cases, but where there is no empirical evidence behind the generalization to unstudied cases. Just because we find that two cases that represent the cross-case diversity of a population are similar at the within-case level, we cannot assume that within-case processes are similar in all other unstudied cases without exploring empirically in other cases whether this assumption holds.

A theory-based strategy involves using existing theoretical knowledge to categorize cases into most likely and typical, or least likely and typical cases (e.g. Eckstein, 1975; Gerring, and Seawright, 2007; Levy, 2008). Here lessons from one case are then used to update our confidence in the validity of an abstract theory across a population of cases based on how it relates on theoretical conditions. For instance, a least likely case that a priori is expected to not fit a theory based on existing knowledge can be selected. If the findings from the least likely case support the theory, a strong generalization is then made that what worked 'here' will also work 'everywhere' (aka a 'Sinatra' generalizing inference). Yet this generalization is based on theoretical speculation about the nature of the population (i.e. some cases are expected to be less likely than others). As with the distribution-based case selection strategy, we still have no actual empirical evidence supporting the generalizing claim. Instead, we are merely assuming without any empirical evidence suggesting that it actually works anywhere else.[8]

[8] - Another strategy is to study all cases, meaning no generalizations to non-studied case are being made. Instead, by studying all cases, the abstract theory is merely a summary of the processual commonalities shared by all cases. However, this still runs into the question of how we can draw processual lessons from particular cases and compare them to see whether it works in similar ways.

Most work on process tracing eludes the question of how we can establish similarity through processual comparisons of two or more cases. Bennett and Checkel (2014:13) admit that 'generalization can be problematic', but offer no guidance for how processes can actually be compared to assess similarity. Beach and Pedersen (2019) do highlight the risk of generalizing processual claims without evidence from other cases, but only provide a 'snowballing outwards' framework for selecting more cases, but they do not tell us how we should assess processual similarity. Bennett (2022) discussing contingent generalization from case studies, but provides no guidance for assessing processual similarity. Providing scholars with methodological guidance that enable evidence-based processual generalizations to be made is the contribution of this article.

2.1. The challenge illustrated

In our own work, we have engaged in in-depth case studies of many of the major crises since 2010 and responses they have triggered, including: banking crisis and Banking Union (Nielsen and Smeets, 2018), eurocrisis and the Fiscal Compact (Smeets and Beach, 2020a), eurocrisis and ESM (Smeets, Jaschke and Beach, 2019), migration crisis and the EU-Turkey deal (Smeets and Beach, 2020b), COVID-19 crisis and the Recovery Fund (Smeets and Beach, 2022), and the British membership crisis and re-negotiation (Beach and Smeets, 2020c).

Given space constraints, each of the articles has a 'stand-alone' character. However, in most of the articles, we do engage in some speculation about whether the processual lessons about how crisis management by institutional networks worked might be 'more broadly' applicable.[9] For example, in our article on the negotiation of the EU-Turkey deal, we write that, '

> 'Obviously, we do not expect to find the same network operating in exactly the same fashion with exactly the same effect in other dossiers, but we do expect to find similar elements (bridging, linking, shielding, laying out tracks, creative fixes) in other cases, in which they operate in a similar fashion and with similar effects. For instance, there are notable

[9] - In some of the articles, we were prodded to do so by reviewers who were skeptical about the utility of only making claims to the studied case.

> similarities with other major reform dossiers, like the British negotiations of 2015-2016 (Beach and Smeets, 2018)...' (Smeets and Beach, 2020b: 147).

Here we make clear that we do not expect the same processes to be operative, but using a soft comparison with our case study of the British re-negotiations, we suggest that similar processual dynamics are at work in the two cases, and therefore might be present in 'other dossiers' (i.e. more generally across cases).

In our analysis of the EU Recovery Fund in the COVID-19 crisis, we engage in three pages of processual comparison, although we lack a clear set of parameters for the comparison. Finally, we engaged in a comparative analysis that attempted to draw out 'general' lessons for how the EU manages crisis post-Lisbon from across the case studies on three parameters – one of which deals with the process (Smeets and Beach, 2020c: 1144). However, as with the 'Failing Forward' special issue (see above), our processual comparison lacks clarity on how we actually detect processual similarity based on the empirics in the different cases, and to what other cases our processual claims might be expected to hold.

3. Learning from cases: an iterated framework for evidence-based generalization

How can we learn from the findings of particular empirical case studies? How can we establish that similar processes are operative across two or more cases? How can we avoid mere theoretical speculation about how things work in other cases, but instead engage in evidence-backed generalizations without having to study all cases?

In this section we develop a set of methodological guidelines that enable more informed processual generalizations. Table 1 describes the four steps in our guidelines. In this section we walk through each step of the guidelines, illustrated based on what our own research program on EU crisis decision-making would have looked like if we had followed the guidelines.

1. Abstraction: select initial case on cause/outcome and develop a process theory based on empirics
2. Identify a target population to which process theory plausibly might work (delineate scope conditions of process theory)
3. Particularize and compare: select another case that is similar on cause/outcome/scope and particularize process theory. Engage in processual comparison by assessing functional equivalences across the two cases. Revise process theory and scope conditions depending on findings.
4. Iterate steps 2-3 to explore bounds of target population (scope conditions for process) using 'snowballing outwards' case selection strategy (ever more diverse cases)

Table 1 – The four step procedure for making evidence-based processual generalizations.

Step 1 - abstraction – process theorization and moving from the particular to the 'general'

Before a process theory can be developed, a case has to be selected. If some form of comparative analysis has been undertaken before the case study, the cause(s) and outcomes are already defined and cases are scored on them. This means that cases are defined as 'cases

of' a cause-outcome relationship as it manifests itself in the cross-case analysis. This comparative data can be used to select a case where the cause(s), outcome and any already identified scope conditions are present in the selected case. If the comparative analysis was done using a set-theoretical method such as Qualitative Comparative Analysis, a positive case would be selected for the initial case study based on the logic that we can only learn about processes linking causes and outcomes by tracing them in negative cases in which the cause(s) that could trigger a process, and/or the outcome that might have been produced by are not present.[10] Of course, the cross-case comparison tells us nothing about whether similar processes work in similar ways because we have not undertaken any within-case analysis that would enable us to say anything about processual dynamics in cases.

An equally valid case selection strategy is to select an interesting empirical case. For example, we might select the EU response to the Covid-19 crisis as a 'case', but where we have not yet identified what it is a 'case of', and what it is about the case that is of theoretical interest (i.e. we do not know exactly what the cause and outcome are). Through abductive back-and-forth 'casing' (Soss, 2018), in which we engage in a dialogue between existing theories and the empirics of a case, we would gradually choose to focus on a particular cause (or set of causes) and outcome and then start to unpack a process theory for how they were linked in the case.

Once a case has been selected, we then engage in process tracing to develop a process theory. Theorizing a causal process is more than depicting actions as *events* in a chronological narrative, in which temporal succession can be conflated with a process being causal (Sayer, 2000: 141). Instead, a process theory describes in more abstract terms the key actors, the activities they are performing in their interactions with each other that link a cause (or causes) with an outcome. Further, instead of detailing every set of interactions between a large number of different actors in a process, a good process theory focuses only on key episodes, defined as interactions between actors that are crucial from a causal perspective,

[10] - Deviant cases where processes 'broke down' and the outcome did not manifest itself can be relevant to study, but only after we have understood how things work in the typical, positive cases. See Beach and Pedersen, 2018 and 2019 for more.

in which we have theoretical reasons to expect that the process might have played out differently if different actions had been taken (Steel, 2008: 88-92).[11]

Actors can be both individual (e.g. President of the European Commission) or collective (e.g. the German government). The only condition for theorizing collective actors is that they have properties and orientations that enable them to *do things* that can impact other actors in a process, and that their activities as a collective cannot be reduced to the sum of the individuals. Activities are what actors do in a process, and are what binds them to other actors in a causal process. To be part of a process, the activities of one actor have to trigger an action in response from other actors. This means that activities are understood in *relational* terms.

However, merely theorizing *what* activities actors are doing is not enough to understand *why* the actions of one actor led other actors to do things. To do this requires that the process theory makes as explicit as possible what Cartwright and Hardie (2012: 22) term the *causal principles* that explain why a given activity might plausibly lead another actor to do something. Causal principles take the form of 'because' clauses when theorizing key episodes. For example, we might theorize that actor A engages in a shaming speech act (activity) directed towards actor B because they perceive the proposal of actor B clashes with shared norms for appropriate behavior. Here the causal principle would be the norm clash motivating the activity. A key benefit of making causal principles explicit is that we can use them as a key metric when assessing the functional equivalence of actors and activities in key episodes (see below for how to do this).

The degree of granularity of a theorized process depends on the research goals. If the research goal is to gain a deeper understanding of a very limited and homogenous population of cases, the theorization of actors and activities will include more actors who will be defined with more attributes, and more episodes of interaction will be unpacked. In other

[11] - Key episodes are similar to critical junctures. We prefer to use the term key episode as it is not tightly bound to a particular theoretical perspective such as historical institutionalist theorization.

circumstances, a simpler process theory can be used to understand process commonalities across a larger number of more heterogeneous cases. However, here the process theory still needs to provide enough detail that it enables assessment of processual similarities and differences. A excessively simplified, 'one-size-fits-all' type of claim tells us next to nothing about how things worked in a case. For instance, a simple one-liner like 'intergovernmental coordination' tells us nothing about who is doing what and why in any given case, meaning we cannot detect whether similar processes are at play in different cases. A good process theory should at the minimum provide us with information about what actors are relevant, what they are doing and why their activities are linked together in key episodes linking a cause and outcome together in a causal sense so that we can detect similarity and differences.

Step 2 - identifying a target population

After a process theory is developed from a case, it is still only a *theoretical* generalization. In other words, we have evidence from the studied case for how it works 'here', but we can only speculate that it works in similar ways somewhere else. Even if our cross-case analysis tells us that the studied case is very similar to a set of other cases (a target population), without additional within-case analyses and processual comparisons, we do not know whether it worked in similar ways in any other case other than the initially studied.

Before we can assess whether similar processes are operative, we have to identify a potential target population. This involves making explicit the scope conditions within which we might plausibly expect the cause and outcome to be linked by the theorized process. Making the activities and causal principles linking them with responses by other actors explicit sheds light on contextual (aka scope) conditions required for the process to operate. For instance, in our studies of crisis responses we have found that the European Council informally delegates certain tasks like drafting potential solutions to an institutional network because they are dependent on the provision of qualified solutions to the problems causing the crisis provided by an actor that is not one of their own. Dependence on the supply of potential solutions from institutional experts is therefore a scope condition that has to be present for the institutional network process to be operative.

Depending on how causally complex the phenomena is, the potential population to which we might generalize is larger or smaller (Rohlfing, 2012: 204-11; Woolcock, 2022). Very complex processes will be very sensitive to context, and vice versa (Beach and Pedersen, 2019; Woolcock, 2022). If we mapped the population using a cross-case analysis, we have concrete identifiable cases that can be selected. If we do not have this data, we can still proceed by searching for another case that is relatively similar on scores of the cause(s), outcome and scope conditions.

In our studies, the target population was defined as cases where the causes 'crisis' and 'need for EU action' being present. Additionally, we only included post-Lisbon cases because this reform strengthened the European Council by replacing the rotating Presidency with a permanent President. A final condition we bounded our target population on was whether the issue was 'Chefsache', i.e. an issue was so sensitive that it required close involvement of the European Council.[12] Our target population is depicted in the right-side of figure 1.

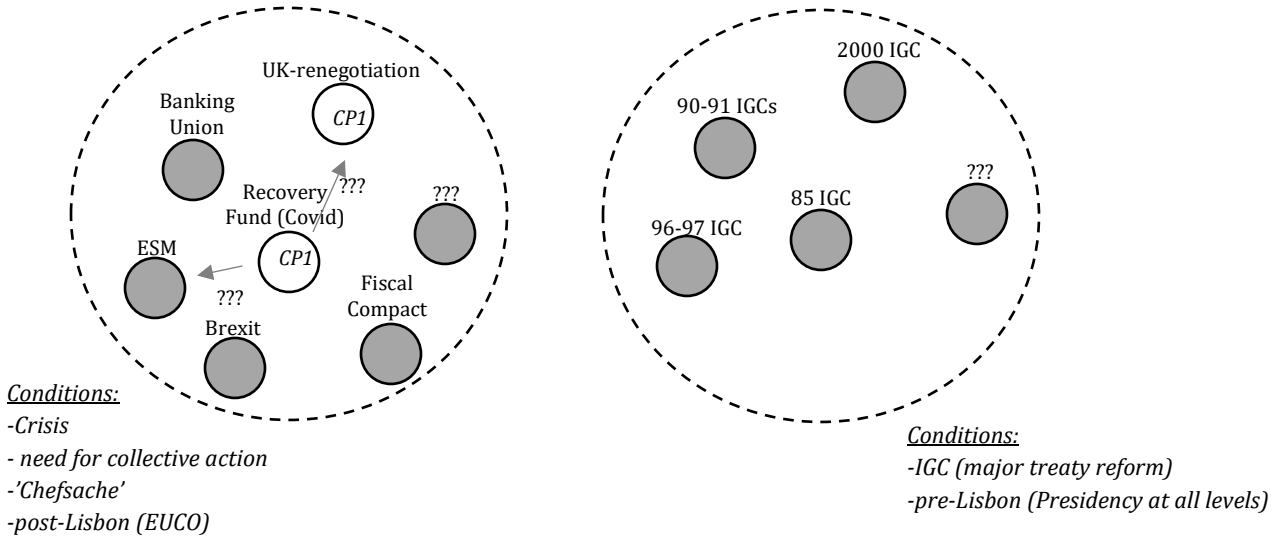

Figure 1 – Bounding target populations and selecting cases

Note : cases are depicted as circles (white = studied with case study, grey = not studied)

[12] - Other potentially relevant scope conditions included whether negotiations were intra-EU or included external parties (e.g. the EU-Turkey deal), and the form of agreement (e.g. a legislative package instead of intergovernmental agreements).

We did not attempt to investigate whether similar processes were operative in other cases of major reform pre-Lisbon because many of these cases were either non-crisis, and/or involved major Intergovernmental Conferences (IGC's) that had very particular institutional dynamics (see Beach, 2005). The target population for our studies was therefore limited to a discrete set of crisis negotiations that have already occurred, as well as potential future cases.

Step 3 – Particularizing and the comparative processual assessment of functional equivalence
A second case from the identified target population is then selected based on based on similar scores on the cause(s), outcome and any identified scope conditions.[13] The purpose of the second case study is not to 'test' whether the process theory is valid per se, as the first case study provided evidence that it was valid in that case. Instead, the purpose is to explore the external validity of the processual claim.

In order to do this, the process theory needs to be 'particularized' in the second case. This involves translating the theoretical abstractions of actors, activities and linkages in key episodes into case-specific empirics. The critical move is to deploy extensive case knowledge of both cases to assess whether the actors and the activities they performed in their interactions during key episodes played causally similar roles, i.e. assessing functional equivalences.

Establishing functional equivalence has a qualitative character, building on detailed empirical case knowledge of the two cases of how things typically work and how they worked in the cases being studied. For instance, in our work on EU crises, one key episode we theorized was that expert institutional actors provide draft texts for negotiations that lay out the tracks *because* the Heads trust them and *because* the Heads are dependent on the expert for translating their ideas into workable solutions. In the British re-negotiation case,

[13] - The second case does not need to be independent from the initial case, as the core causal inferences are made *within* each case. In our research, how the EU managed crisis in early cases (e.g. the ESM) shaped how they responded to subsequent cases – meaning the cases are strictly speaking not independent of each other as the lessons of one spilled into subsequent cases.

the actor providing the formative texts was the Council Secretariat Legal Service, and the activity involved producing a skeleton draft text early in the process that laid out the tracks for further negotiations (Beach and Smeets, 2020). In the case of the EU Recovery Fund, a functionally similar (but NOT identical) role was played by budget experts from several Commission DG's, who developed a legal recipe for how the EU could collectively borrow money without changing the treaties or making the scheme permanent (Smeets and Beach, 2022). For both cases, these actions were taking part during the agenda-setting phase early in the negotiations (one key episode).

We can also use the findings of the processual comparison to refine a given process theory. Are there irrelevant differences that can be removed/modified in the process theory? Does a causal principle behind a given linkage need to be made more explicit. Were there additional causes that were initially unknown, but that were detected after we had engaged in extensive probing in both cases?

If we find processual similarity across the two cases, we can return to the theoretical level with an evidence-based generalization for how it works 'here' (case 1) and 'there' (case 2) (see figure 2, below). If we find different processes – either for the whole process or for key episodes – what can account for differences? Are there important differences in the scope conditions that plausibly produced different processes? For example, there might be different process dynamics between actors when the negotiations involve intra-EU discussions versus an us-them type debate, as in the British membership crisis. The new knowledge gained about scope conditions can then be used to split the target population into two subpopulations within which different processes are operative.

If we had used our iterated procedure, we might have selected COVID-19 crisis and the EU Recovery Fund as an initial case. A process theory abstracted from the case is depicted in the first row of table 2. The top row depicts the key episodes (e.g. laying out tracks, tasking, etc), defined as the moments in the process where an institutional network interacts with either the European Council (EUCO) or machine room actors (Council) that are critical in terms of how the process was managed and during which key functions such as drafting potential

solutions were performed. Below this are the actors and activities that provide the causal linkages in the process. Note that for some functions, they might be performed at different points throughout the process (e.g. laying out tracks often is repeated at other key episodes later in the process). The process theory only details one causal linkage between crisis and reforms, and therefore is not an exhaustive (aka sufficient) explanation. Instead, it should be understood as one part of a typically quite complex, multi-factor causal explanation. However, the process theory here highlights one set of interactions that provide an important contribution to the outcome.

		laying out tracks	tasking	Negotiations	Remandating	Endgame	
EU Recovery Fund case (2020)	**Crisis** (demand for collective EU action) + **Chefsache**	**Institutions** *identify and draft* potential solutions for EUCO	**EUCO**: 1) *provides* broad parameters for MR debates, 2) **Heads** *informally delegates* to institutional network	**Institutional network** *provides* process management: 1) *shield* 2) *bridge* institutional divides 3) *inform* CR and renew support (*linkage*)	**Heads** *supply* political leadership, 1) *provide* support for MR process 2) *set* deliverables deadlines	**Heads** *resolve* final outstanding issues (political leadership), *assisted* by **institutions** that *find creative fixes*	**Final reform** EU Recovery Fund that is much more ambitious than governments left to themselves would have agreed upon
UK renegotiation case (2015-16)	**Crisis** (demand for collective EU action) + **Chefsache**	-similar dynamics in episode, but actor different (Council Secretariat provides initial drafts)	-less EUCO guidance -more explicit tasking (EU Task Force (EUTF) created to 'assist' EUCO	-drafting in EUTF, with input from UK -input provided by 'Sherpas' prior to EUCO	-similar dynamics	-similar dynamics at EUCO summit	**Final reform** arguably more accommodating towards UK than EU governments would have agreed upon themselves

Table 2 – Processual comparison across two cases of crisis.

We could then have selected a second case of crisis->reform to assess whether similar processes were at play in any other cases. For example, we might have selected the UK re-negotiation case, which was a crisis triggered by Cameron's call for a referendum on revised

terms of membership for the UK. The two cases differed considerably in terms of which actors were performing different functions (e.g. the role of the Council Secretariat in supplying drafts was much more prominent in the early stages of the UK re-negotiation), and how the negotiations were structured differed. In the Covid-19 case, there were repeated EUCO meetings that set the broad parameters for debate, after which institutions shaped the form of solutions that were then discussed at the highest level. In contrast, there were fewer EUCO discussions in the UK re-negotiation case, and more extensive 'tasking' of expert actors to flesh out a fair deal that responded to UK demands within the realm of the possible (i.e. no treaty change). However, when compared in terms of functional equivalence at key episodes, we see a similar pattern of institutions developing potential solutions quite early in the process (laying out tracks), informal mandating by the EUCO of different process management and drafting functions to an institutional network with varying composition (tasking), and similar activities performed by institutional actors in providing creative fixes in the end game. The two case studies shed light on the scope conditions required for the process to work. For instance, the central role of the EUCO is only required in 'Chefsache' type cases that are so sensitive that the Heads of State and Government have to be closely involved in reform negotiations.

After this comparison, we have evidence of (relatively) similar processual dynamics in the two studied cases. Given that there are other cases that share the cause(s), outcomes and scope conditions (see figure 1), we would be able to make a very cautious processual generalization in which we have made it slightly more plausible that similar processes might be operative in the other cases as well. If we wanted to be able to make stronger evidence-based processual generalizations, we need to explore whether similar dynamics were also present in other cases, which is the last step in our guidelines.

Step 4 – iterate as required to increase confidence in processual generalizations

If we want to strengthen our ability to make evidence-based generalizations beyond the two initially studied cases, we can deploy a more extensive phase of comparative processual analysis, selecting ever-more diverse cases to explore empirically the bounds of valid processual generalizations, asking in effect how far can it travel.

Table 3 depicts a range of parameters for determining when relatively strong processual generalizations in the form of 'process X is probably also operative within the rest of the target population'. While we will never be highly confident in similar processes being present in other, unstudied cases unless we actually open the hood and assess them, we can increase our confidence after we have explored more cases.

Higher confidence processual generalizations (probably)	**Lower confident processual generalizations (might)**
more abstract process theory	more specified process theory (more actors, activities, key episodes)
low causal complexity (homogeneity)	high causal complexity (heterogeneity)
process less sensitive to context (few scope conditions)	process highly sensitive to context (many scope conditions)
high ratio of studied/unstudied cases	low ratio of studied/unstudied cases
diverse cases within target population assessed (similarity found)	not diverse selection of cases

Table 3 – When are we justified in making strong processual generalizations?

For example, if there are three cases that are very similar on cause(s), outcome and scope conditions, and we have studied cases 1 and 2 and found similar processes, we can infer that a similar process is *probably* also present in case 3 (high confidence processual generalization based on evidence from cases 1 and 2). However, if we have a target population of seven cases that differ on a range of potentially relevant conditions, and we have only studied cases 1 and 2, relatively low confidence processual generalizations to the

other cases are possible. If the cause->causal process->outcome relationship is very complex, only weaker processual generalizations to other cases should be made.

Note that selecting ever more diverse cases in a target population is not a 'mechanical' approach, but instead choices should be based on theoretical/empirical knowledge of the phenomena being studied when selecting additional cases. For instance, we might be particularly concerned that our process theory of consensual decision-making might not hold for distributive issues. We would therefore explore whether similar consensual processes found operative in non-distributive issues also were present in distributive issues.

For practical reasons, we suggest that additional processual comparisons in the extensive phase use what can be termed PT 'light', in which the focus is on assessing whether one or more key empirical signatures from one or more key episodes is present or not in any given case (Steel 2008: 88-92; Beach and Pedersen, 2019:133–45). Cases can then be selected that are ever-more diverse cases to explore whether similar processes are operative or not. If we find similar processes, we can eliminate a scope condition (e.g. distributive/non-distributive), and thereby refine the target population further.

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 subpopulations.

4. Conclusions

This article put forward guidelines for drawing theoretical lessons from particular cases, identifying relevant target populations, and then assessing empirically whether similar processes are operative in other cases in order to enable more 'general' evidence-backed claims to be made about how things work within a bounded population of cases. We used our own research on EU crisis management and informal delegation by the European Council to illustrate how it can work in practice.

Figure 2 (next page) summarizes the four step guidelines for making evidence-based processual generalizations. The procedure involves first the abstraction of processual lessons (aka theorization) from a particular case, followed by an identification of a plausible target population of other cases within which the process might work in similar ways. However, at this stage this is only a *theoretical* generalization, and moving towards evidence-based generalizations requires additional case studies that assess whether the process worked in *functionally equivalent* ways in other cases (or not) that gradually increase the evidence behind our claims about the scope of the population of cases within which our processual claims hold. At the process level, functional equivalences deal with who is doing things (actors), what they are doing and how they are linked in a causal sense (actions and linkages) in key episodes in the causal process linking a cause triggering it and an outcome.

Comparing cases at the process level results in better delineation of target populations. By tracing how things work in different cases, we shed more light on the scope within which our processual generalizations can be expected to be applicable. Additionally, by assessing functional equivalences in how processes work across two or more cases, we also produce better processual theories because we are able to abstract away case-specific linkages through comparing.

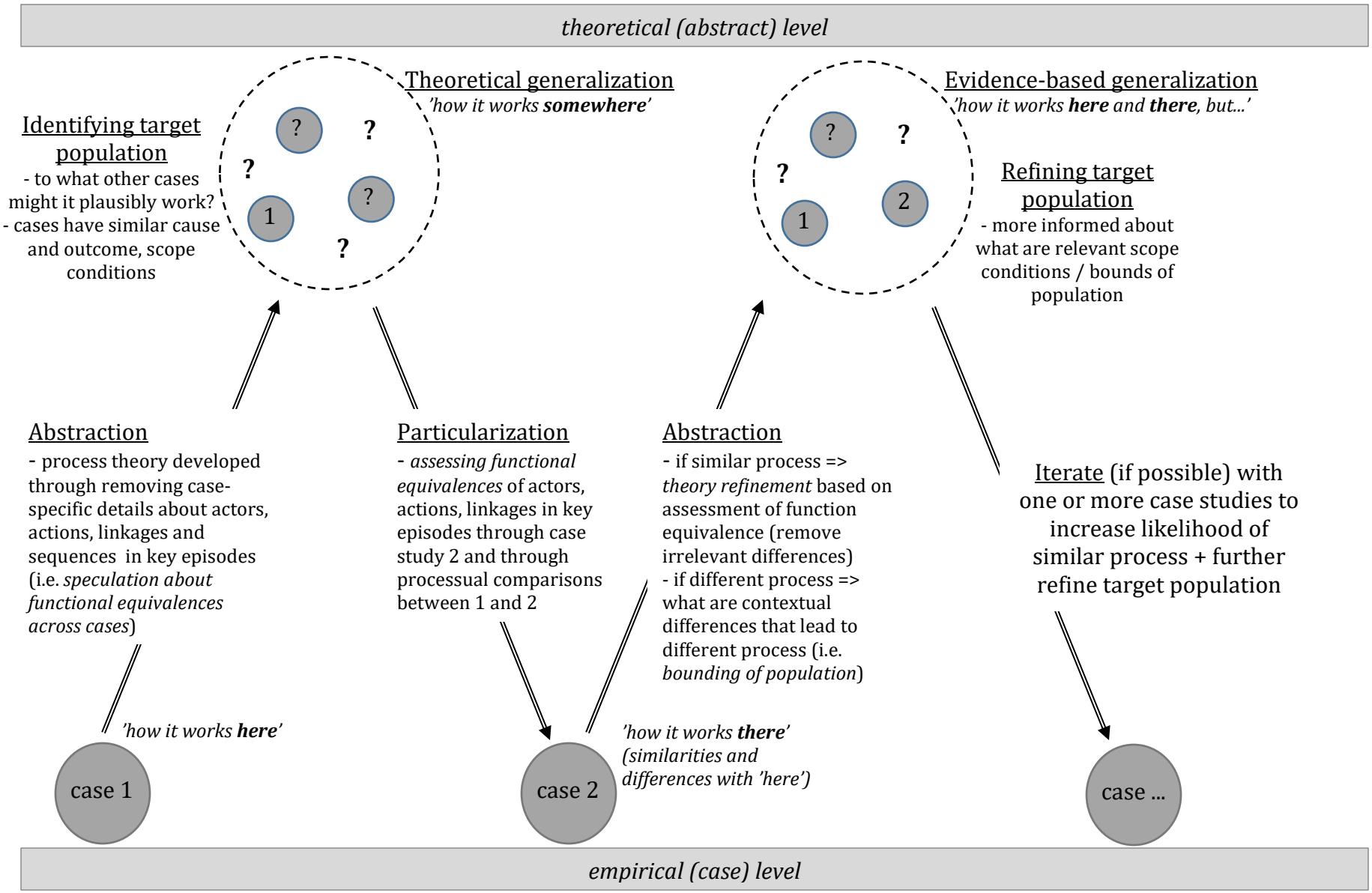

Figure 2- An iterated design for developing evidence-based generalization using process tracing case studies.

References

Beach, D.. 2005. *The dynamics of European integration: when and why EU institutions matter.* Houndmills: Palgrave Macmillan.

Beach, Derek and Rasmus Brun Pedersen. 2018. Selecting appropriate cases when tracing causal mechanisms. *Sociological Methods & Research*, 47(4):837-871.

Beach, D. and R.B. Pedersen. 2019. *Process-tracing – foundations and guidelines.* 2nd Edition. Ann Arbor: University of Michigan Press.

Beach, D. and S. Smeets. 2020. The unseen hands – collaborative instrumental leadership by institutions in the British re-negotiation case. *European Journal of Political Research,* 59(2): 444-464.

Bennett, A., and J. Checkel. 2014. *Process Tracing: From Metaphor to Analytic Tool*. Cambridge: Cambridge University Press.

Bennett, A.. 2022. 'Drawing Contingent Generalizations from Case Studies.', in Widner, Woolcock and Nieto (eds) *The Case for Case Studies: Methods and Applications in International Development.* Cambridge: Cambridge University Press, pp. 62-86.

Cartwright, N. and J. Hardie. 2012. *Evidence-based policy: A practical guide to doing it better*. Oxford University Press.

Eckstein, H.. 1975. Case Study and Theory in Political Science. In *Handbook of Political Science – Strategies of Inquiry Vol. 7,* ed. Fred I. Greenstein and Nelson W. Polsby, 79-138. Reading: Addison-Wesley Publishing Company.

Gerring, J. and J. Seawright. 2007. 'Techniques for Choosing Cases'. In *Case Study Research*. Cambridge: Cambridge University Press, pp. 86–150.

Jones, E., R. D. Kelemen, and S. Meunier. 2016. Failing forward? The euro crisis and the incomplete nature of European integration. *Comparative Political Studies*, 49(7), 1010–1034.

Jones, E., R. D. Kelemen and S. Meunier. 2021. Failing forward? Crises and patterns of European integration. *Journal of European Public Policy*, 28:10, 1519-1536.

Levy, J.. 2008. Case Studies: Types, Designs, and Logics of Inference. *Conflict Management and Peace Science,* 25 (1): 1–18.

Nielsen, B., and S. Smeets. 2018. The role of the EU institutions in establishing the banking union. Collaborative leadership in the EMU reform process. *Journal of European Public Policy,* 25(9): 1233-1256.

Przeworski, A. and H. Teune. 1966. Equivalence in Cross-national Research. *Public Opinion Quarterly,* 30(4): 551-568.

Rohlfing, I.. 2012. *Case Studies and Causal Inference: An Integrative Framework*. Houndmills: Palgrave Macmillan.

Sartori, G.. 1970. Concept Misformation in Comparative Politics. *American Political Science Review,* 64(4): 1033-1053.

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

Smeets, S., A. Jaschke, and D. Beach. 2019. The role of the EU institutions in establishing the ESM. Institutional leadership under a veil of intergovernmentalism. *Journal of Common Market Studies, 57*(4), 675-691.

Smeets, S., and D. Beach. 2020a. Political and instrumental leadership in major EU reforms. The role and influence of the EU institutions in setting-up the Fiscal Compact. *Journal of European Public Policy*, 1-19. doi: 10.1080/13501763.2019.1572211.

Smeets, S., and D. Beach. 2020b. When success is an orphan. Informal institutional governance and the EU Turkey deal. *West European Politics* 43(1): 129-158.

Smeets, S., and D. Beach. 2020c. Intergovernmentalism and its implications. New institutional leadership in major EU reforms. *Journal of European Public Policy*, *27*(8), 1137–1156.

Soss, Joe. 2018. On Casing a Study versus Studying a Case. *Qualitative and Multi-Method Research*, 16(1): 1. doi.org/10.5281/zenodo.2562167

Steel, Daniel. 2008. *Across the Boundaries: Extrapolation in Biology and Social Science*. Oxford: Oxford University Press.

Tavory, I. and Timmermans, S.. 2014. *Abductive Analysis: Theorizing Qualitative Research.* Chicago: University of Chicago Press.

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.