AARHUS UNIVERSITY

Coversheet

This is the accepted manuscript (post-print version) of the article.

Contentwise, the post-print version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication

Please cite the final published version:

Møller, J., & Skaaning, S-E. (2017). Explanatory Typologies as a Nested Strategy of Inquiry: Combining Cross-Case and Within-Case Analysis. *Sociological Methods & Research*, 46(4), 1018-1048. DOI: 10.1177/0049124115613778

Publication metadata

Title:	Explanatory Typologies as a Nested Strategy of Inquiry: Combining Cross-Case and Within-Case Analysis
Author(s):	Møller, J., & Skaaning, S-E.
Journal:	<i>Sociological Methods & Research</i> , 46(4), 1018-1048
DOI/Link:	http://dx.doi.org/10.1177/0049124115613778
Document version:	Accepted manuscript (post-print)

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Explanatory Typologies as a Nested Strategy of Inquiry:

Combining Cross-Case and Within-Case Analysis

Jørgen Møller

Professor, PhD

Department of Political Science, Aarhus University

Bartholins Allé 7

8000 Aarhus C, Denmark

E-mail: jm@ps.au.dk

&

Svend-Erik Skaaning

Professor, PhD

Department of Political Science, Aarhus University

Bartholins Allé 7

8000 Aarhus C, Denmark

E-mail: skaaning@ps.au.dk

Explanatory Typologies as a Nested Strategy of Inquiry: Combining Cross-Case and Within-Case Analysis

Abstract

Explanatory typologies have recently experienced a renaissance as a research strategy for constructing and assessing causal explanations. However, both the new methodological works on explanatory typologies and the way such typologies have been used in practice have been affected by two shortcomings. First, no elaborate procedures for assessing the general explanatory power of a typological theory on the cross-case level have been devised. Second, rigorous selection procedures for within-case analysis are lacking. Against this background, we introduce a systematic measure that helps researchers assess the explanatory power on the cross-case level, first, within the scope set by a particular typological theory and, second, by investigating the transferability of the theory beyond these scope conditions via an increase of the number of cases. Drawing on recent methodological works on nested analysis, we show how researchers can identify key cases for process tracing based on the cross-case explanatory fit of the typological theory. We illustrate the purchase of our procedures by revisiting seminal studies from the field of comparative historical analysis.

Keywords:

Explanatory typologies, case selection, nested analysis, comparative historical analysis

Introduction

The use of typologies has always figured prominently in social science. In some research fields, they have been virtually ubiquitous. For instance, practically all major contributions to the field of comparative historical analysis (Mahoney and Rueschemeyer 2003) – from Moore ([1966] 1991) to Mahoney (2010) – can be understood in terms of explanatory typologies. However, typologies have been criticized from two different vantage points. Quantitative scholars have argued that they are inferior to more sophisticated techniques of data analysis such as factor analysis and cluster analysis (Bailey 1994; Ahlquist and Breunig 2012), and qualitative scholars have berated the lack of attention to case-based micro-processes on which the causal claims of typologies are based (Tilly 1997; Capoccia and Ziblatt 2010).

Nonetheless, typologies have recently experienced a renaissance in social science, both descriptively and as a research strategy for constructing and assessing causal explanations (Elman 2005; George and Bennett 2005; Bennett and Elman 2006; Smith 2002; Collier, LaPorte, and Seawright 2012). In this paper, we focus on the second purpose, what has interchangeably been termed typological theory and explanatory typologies.¹ Our objective is to develop systematic procedures for this strategy of causal assessment, which addresses both the quantitative and the qualitative concerns about typologies. We follow Elman's (2005:298) notion that "[a]n explanatory typology is based on explicitly stated pre-existing theory" and that "filling in the cells requires working through the logical implications of the theory: given its posited causal relationships, what particular outcomes are associated with different combinations of the theory's variables?" An explanatory typology is thus a product of deductive theorizing and it entails the explicit formulation of hypotheses about patterns in the case distribution within a particular property space (see also George and Bennett 2005:235). What is less clear from extant scholarship is whether the typology

serves to test this theory or whether it only purports to serve as a heuristic device for identifying instances that are relevant for within-case analysis.

For two reasons, we come down in favor of assessing the proposed hypotheses based on the cross-case distribution. First, most scholars see explanatory typologies as a nested strategy of inquiry that, perforce, combines cross-case and within-case analysis (see George and Bennett 2005; Bennett and Elman 2006). An established premise in the literature on case selection in nested analysis is that the cross-case robustness of a theory has implications for which cases should be selected for in-depth study (e.g., Lieberman 2005; Rohlfing 2008). Second, a cross-case test is necessary to adjudicate between competing typological theories. Scholars routinely criticize prior research for not adequately elucidating the relevant empirical variation when they introduce new explanations that take the form of explanatory typologies (e.g., Ertman 1997; more on this below). They thus claim that their new explanation has a better empirical fit. In the absence of systematic tools for assessing the empirical fit of different theories, these claims remain unconvincing.

Seen in this light, two problems affect the recent works on the use of typologies (Elman 2005; George and Bennett 2005; Bennett and Elman 2006; Collier et al. 2012). First, these discussions have not included systematic guidelines about how to measure of the general explanatory power of a typological theory based on the case distribution within the property space. Second, while most work on explanatory typologies has concerned how to select cases within the property space for process-tracing (Bennett and Elman 2006:467; George and Bennett 2005:251-53), the proposed guidelines for case selection have not absorbed recent debates about case selection in nested analyses (e.g. Lieberman 2005; Seawright and Gerring 2008; Fearon and Laitin 2008; Schneider and Rohlfing 2013). As already alluded to, the two problems are interlinked. In the absence of prior considerations about the extent to which an explanatory model accounts for cross-case patterns, we do not know which cases to select for within-case analysis (Rohlfing 2008).

Recall in this connection that recent scholarship has convincingly stressed the enduring value of controlled comparisons (Tarrow 2010; Slater & Ziblatt 2013). It is exactly by nesting case studies within this foil that typological theory becomes a potentially powerful tool. In what follows, we propose systematic procedures for, first, the assessment of explanatory power on the cross-case level and, second, subsequent case selection for within-case analysis. Regarding the cross-case level, we argue that scholars using typological theory should calculate the fit between the theoretical expectations and the case distribution and compare it to the (hypothetical) fit of a random case distribution. If a strong fit between the theoretical expectations and the case distribution is established, and if it is not possible to carry out process tracing of all cases, we argue that scholars should pay special attention to two kinds of cases. First, cases representing each of the explanatory factors, or combinations of explanatory factors, that are sufficient for the outcomes of interest. This can be understood as an attempt to use new within-case evidence to test whether the causal mechanisms on which the typological theory is based can be corroborated. Second, deviant cases should also be analyzed in depth to scan for additional explanatory factors that can account for the misfits.

If, on the contrary, the case distribution on the cross-case level does not corroborate the theory, scholars can use process tracing to facilitate theory building by focusing on cases exhibiting clear membership in the outcomes of interest. To search for new explanations, these cases can be contrasted with cases that have different outcomes but score fairly similarly on explanatory factors stressed by extant theories. Finally, if several cases fulfil the proposed criteria, or if it is untenable for researchers to carry out process tracing in all the cases suggested by the guidelines, more pragmatic reasons based on the number of cases covered by the different configurations, the perceived paradigmatic status of cases in a particular research field, “local knowledge”, and access to relevant sources can be taken into consideration.

We proceed by specifying our procedures for calculating the fit between theory and case distribution and the guidelines for case selection that follows. Next we illustrate these procedures by revisiting a number of influential analyses from comparative historical analysis (Mahoney and Rueschemeyer 2003), some of which are discussed in the paper, and some of which are treated in an online appendix. The final section concludes.

Assessing cross-case explanatory power

It follows from our initial discussion of explanatory typologies that the first step in testing a typological theory is to assess its explanatory power based on the case distribution *tout court*. More particularly, what we need is a way to evaluate the fit of theories that, within the confines of a deductive property space, propose that certain conditions, or combinations of conditions, are necessary and/or sufficient for an outcome of interest (see Elman 2005: fn. 15; Schneider and Rohlfing 2013: fn. 4). The attempt that comes closest to doing so in prior research is Ragin's (2000) discussion of the quasi-sufficiency of causal combinations.

Ragin's focus differs somewhat from ours as he is more occupied with the extent to which the cases are consistent with the sufficient (combinations of) conditions identified in the *empirical* analysis whereas we are more interested in the extent to which the cases fall in the *theoretically predicted* configurations of an explanatory typology. Nonetheless – with some important alternations – Ragin's procedures can be used to assess the fit of typological theories. More particularly, Ragin (2000: 109-16) proposed that scholars should chose a benchmark (null hypothesis) proportion of cases that should fulfill the criteria and then assess if the actual proportion of cases fulfilling these criteria are significantly higher than the benchmark. Moreover, he suggested a series of different benchmarks (more often than not = .50; usually = .65; almost always = .80), and basically argues that it is a theoretical question which one should be chosen.²

We are uncomfortable about setting benchmarks in such an arbitrary way and below propose a solution that fixes the benchmark not in terms of a particular proportion but as a function of how broad the predictions of the typological theory in question are. This is also necessary for adjudicating between different typological theories in a convincing way. A neat way of illustrating why Ragin's guidelines should not, without further ado, be transferred to assessing the fit of an explanatory typology is by revisiting Mahoney's (2003a: 345-346) assessment of Ertman's (1997) analysis of state-regime development.³ Mahoney first shows that Ertman (1997) eliminates rival propositions using a procedure based on considerations about explanatory power. Ertman argues that the extant explanations of scholars such as Anderson (1974), Tilly (1992), and Downing (1992) are contradicted by at least four out of his 13 cases. Ertman's own theory accounts for 11 out of 13 cases and, therefore, has a competitive edge. Mahoney calculates the respective proportions of cases that corroborate the theory (.86 in Ertman's case as compared with a maximum of .71 for the competing explanations) and demonstrates that only Ertman's fit is significantly⁴ higher than a benchmark proportion of 65 percent correctly predicted cases.

However, Mahoney's direct comparison between Ertman's theory and competing theories is unable to appreciate a fundamental point with respect to the use of explanatory typologies: The broader the expectations, that is, the more types are predicted by the theory, the easier it will be to achieve a high fit, all else equal. In other words, both Ertman's implicit test of the different explanations and Mahoney's explicit assessment of this test ignore the scope of the competing explanations. This issue is certainly not trivial with respect to the use of explanatory typologies as some of the prominent theories within comparative historical analysis, including those discussed by Ertman and Mahoney, differ starkly with respect to the proportion of types predicted to contain referent.

We propose that this problem – and the more general absence of any description of a measure of the explanatory power – can be remedied by expanding Mahoney’s assessment in a particular way. When framing Ertman’s analysis in terms of an explanatory typology, it is evident that Mahoney calculates what is normally termed the coefficient of reproducibility (henceforth CR) (see Bailey 1973), which is simply the proportion of cases in a typology classified as expected theoretically. The calculation of the CR can be illustrated using one of the competing theories that Ertman rejects, namely that of Downing (1992).

Downing (1992) sets out to explain why some European countries became instances of what he terms “military-bureaucratic absolutism” whereas medieval constitutionalism survived in others, following the onset of the 16th century military revolution. Downing only includes cases which were characterized by “medieval constitutionalism” before this dynamical factor kicked in, viz., Prussia, France, Poland, England (at two different points in time), Sweden, and the Dutch Republic, the first two of which became absolutist whereas the other four retained the political *status quo ante* (constitutionalism).

To explain this variation, Downing introduces two explanatory factors. The first of these is the intensity of warfare after the military revolution. This variable Downing dichotomizes into high and low. Only high levels of warfare paved the way for the absolutist turn as intense geopolitical pressure required resources which were difficult to extract in the face of constitutionalist barriers. However, as the military revolution in time became an all-out European affair, even England – which was initially protected against the absolutist impulse due to its geographical isolation – eventually became embroiled in the spectacle. High intensity thus changes from an operative variable to a constant in the course of Downing’s analysis. But this is where the second explanatory factor enters. Military-bureaucratic absolutism was only the requisite response if warfare had to be financed by mobilizing the domestic economy. This was the case in Prussia and

France. Where other sources of finance could be found, constitutionalism instead survived. On this basis, the following explanatory typology (with cases) can be created.

[Table 1 about here]

In Table 1, we have shaded the configurations that corroborate Downing's theory, that is, type 1 and types 6-8. The CR of this theory is .86 (6/7). The rule of thumb says that a strong relationship should have a coefficient of reproducibility (CR) above .85. However, the objection about the scope of predicted configurations needs to be countered here. We are equally skeptical about clinging automatically on to, say, a .85-criterion as about arbitrarily setting null hypothesis benchmarks such as .50, .65, and .80. Drawing on work on non-parametric statistical testing based on permutation methods (see Gibbons and Chakraborti 2010), we argue that it makes more sense to assess the actual CR of a given hypothesis with the equivalent CR of a random distribution as a measuring rod. The random CR simply expresses the proportion of types predicted by the theory out of all types in the property space. More particularly, the random CR of the theory in question equals a coin-flipping null hypothesis. In Table 1, the random CR is .5 (4/8) as four types out of the eight occurring are predicted to contain referents. Even though the absolute fit of Downing's theory is high (.86), this frame of reference means that the relative fit is not all that impressive.

We can take the analysis one step further by using the difference between the actual CR and that of the random distribution of cases to calculate whether a given distribution is statistically significant based on the theoretical predictions.⁵ Such a test can help us to take into account how broad the conjectures are. Arguably, a binomial probability test is the most appropriate choice of significance test in our examples because it does not require as strong assumptions about

the number of observations, normal distribution, randomness, independence, and measurement level as conventional parametric tests (see Gibbons and Chakraborti 2010).⁶

Invoking these guidelines, the fit between Downing's theory and the empirical cases, as illustrated in Table 1, is statistically significant only at the .1-level ($p=.054$). This is a higher p -level than that of most of the other analyses replicated in this paper (and in the online appendix), partly because Downing's theory predicts cases to cluster in a higher proportion of the types of property space and partly because his case number is relatively low.

The test is obviously very simple, which can be considered an advantage considering recent skepticism about using more sophisticated tests on samples with a low number of cases (Collier 2014). Moreover, the test provides a systematic response to the problem that more informal assessments of fit based on simply scanning the case distribution in the property space are likely to falter because it is all too easy to ignore the relative fit based on the number of cells in the typology – as done by both Ertman (1997) and Mahoney (2003a).⁷

However, the use of the CR of a random distribution raises a particular caveat. The number of types in a typology is a function of the number of dimensions and the number of classes on each of the dimensions. For instance, a dichotomization of three dimensions produces eight types (2^3); a trichotomization of four dimensions produces 81 types (3^4). Consequently, by adding either dimensions or classes on dimensions which do not alter the main empirical pattern, one can boost the CR vis-à-vis the random CR. It is important to note that if all dimensions and classes are solidly anchored in theory, this problem disappears. Hence, when critically assessing the explanatory power of any typological theory, one should consider carefully whether reductions in the number of types are pertinent. Indeed, any proper use of typologies necessitates reflections about reductions of the property space (see Lazarsfeld and Barton 1951; Elman 2005). Lazarsfeld and Barton (1951)

devised a number of strategies of reduction. Elman (2005) elaborates these nicely, thereby distinguishing between five different ways to reduce the number of types.

The use of reduction can be illustrated by returning to the Downing example illustrated in Table 1 where we have in fact already employed one such strategy. Downing (1992) actually operates not with two but three different classes on the financing of warfare variable. Besides high domestic mobilization he distinguishes between whether warfare could be financed by taxing the commercialized economy (as in England II and the Dutch Republic) or whether it could be financed by extracting tribute from conquered territories (as in the Swedish case). But as the effect of the two latter classes is exactly the same – the preservation of constitutionalism in spite of high levels of warfare – these two classes can be collapsed into one, characterized by the more general existence of alternative sources of finance.

What we employ in the Downing example is the strategy which Elman (2005) terms ‘pragmatic compression’ where continuous cells are collapsed. In this particular instance, the reduction is warranted because the two different classes do not increase the theoretical leverage in any way – but via their existence would have lowered the random CR. Another of the reduction strategies discussed by Elman deserves special attention, namely, “logical compression,” i.e., the deletion of logically impossible types. Obviously, such logically impossible combinations should not be included when calculating the CR of a random distribution; the CR should thus be calculated *after* the completion of logical compression (see also George and Bennett 2005:249). Summing up, our points about reductions of the property space can be summarized in the following general rule: That reduction of the property space is warranted whenever two different classes do not increase the theoretical leverage in any way but lower the random CR.

Case selection

The considerations about the cross-case distribution present the frame of reference for discussing case selection for within-case analysis. One of the repeated recommendations of recent works on explanatory typologies is that they can be used for case selection (Bennett and Elman 2006:467; George and Bennett 2005:251-53). George and Bennett (2005:251-52) propose four research designs based on selecting cases within the property space of an explanatory typology:

... comparing similar or differing cases in the same type; comparing most similar cases in adjacent types with differing outcomes; studying most-likely, least-likely, and crucial cases; comparing least similar cases.

These suggestions may indeed be valuable. However, they provide nothing distinctive about picking cases using typologies. Basically, the problem with this approach to using typologies to pinpoint cases for within-case analysis is that it does not systematically link case selection with the cross-case explanatory power of the typological theory.⁸ We argue that more is needed to achieve a genuine integration of comparative and within-case analysis through typological theory (cf. George and Bennett 2005: Ch. 11).

The only fast and hard case-selection guideline we can establish without any considerations about cross-case patterns is that a purposive case selection (Seawright and Gerring 2008) is preferable for singling out the types of interest (see also Lieberman 2005). However, the purposes of the selection differ depending on whether the explanatory typology is corroborated based on the case distribution. A case study carried out after assessing the cross-case patterns can either serve as theory testing or theory building. If the typological theory is corroborated on the cross-case level, theory testing is the most important purpose of in-depth case studies. If, on the contrary, the theory does not receive much backing on the cross-case level, theory building should

be the most important aim of subsequent case studies (Lieberman 2005; Rohlfing 2008). The two-step approach of calculating the CR of a typological theory and comparing this with the random CR provides a formalized measure of the cross-case strength of an explanatory typology and, hence, indicates what the primary purpose of the case selection should be.

Suppose, first, that the explanatory typology resiliently accounts for the cross-case variation. In this instance, we would first and foremost select what Seawright and Gerring (2008) term “typical cases” to investigate whether the causal mechanisms on which the typological theory is based actually underpin the cross-case regularities (see also Lieberman 2005). This can be understood as an attempt to test the theory against new evidence from the same cases based on within-case logic of inference (Seawright and Gerring 2008:299).

But exactly which cases are typical within the property space of an explanatory typology? Inspired by recent works on case selection for process tracing after QCA (Schneider and Rohlfing 2013; Mikkelsen 2015, Beach and Rohlfing 2015), we can specify the recommendations further. However, before doing so an important difference between explanatory typologies and QCA must be reiterated (see fn. 1 above). In contrast to QCA, explanatory typologies can simultaneously capture multiple outcomes of interest. Indeed, in virtually all of our replications, the aim of the theory in question is to explain both the presence and the absence of a particular outcome or set of outcomes (say, the advent of military-bureaucratic absolutism and the preservation of constitutionalism in the Downing illustration above or the four different combinations of regime forms and state infrastructures in the first Ertman illustration below). This difference means that cases from more configurations should be subjected to process tracing than what follows from guidelines in the recent literature on combining QCA and process tracing, where only the configurations including the “positive” outcome of interest are singled out for within-case analysis (see Schneider and Rohlfing 2013; Beach and Rohlfing 2015; but see also Mikkelsen 2015).

More particularly, with the one exception explained below, we advise scholars to investigate at least one case⁹ that represents each of the sufficient conditions, or combination of conditions, for all the outcomes of interest (see Goertz 2008:11-14; Schneider and Rohlfing 2013). If the theoretical framework predicts equifinality with respect to a certain outcome, all of the predicted configurations make up types containing typical cases. Such cases are normally referred to as “diverse cases” (Seawright and Gerring 2008), but when premised on a typological theory, it makes little sense to distinguish the two kinds of cases (see Rohlfing 2012:70-71).

Illustrating the principles using the Downing example in Table 1, we recommend that at least one case from each of the types 1, 6, and 7 should be selected for process tracing to test if the causal mechanisms on which Downing’s theory is based are to be corroborated by within-case evidence. Cases situated in the last typical configuration, type 8, are less relevant for subsequent process tracing. In the words of George and Bennett (2005:251), “[w]hen an outcome is overdetermined by existing theories and it turns out as expected, it is less likely to be theoretically informative ... Such most-likely cases are usually useful only when a theory unexpectedly fails to explain them” (see also Mahoney and Goerz 2004). Cases in the three other typical configurations are simply more analytically interesting, that is, their potential leverage is higher when it comes to disentangling the relationship between the two conditions and the outcomes in question. Through a selection of cases from types 6 and 7, we can use process tracing to investigate what happens if one of the sufficient conditions of the outcome(s) takes a different value vis-à-vis the configurations, including the alternative outcome, military-bureaucratic absolutism. However, the fact that there are no empirical referents in type 7 means that process tracing of a case in type 8 becomes important – as a next-best option – to investigate the influence of low levels of warfare.¹⁰

When choosing between instances situated in the most relevant types, one can either choose randomly (in effect using a stratified random selection), focus on the cases where “local

knowledge” (language proficiency, contacts, previous research experience, etc.) come in handy (Seawright and Gerring 2008:299), select the cases that tend to provide the best access to reliable sources, or select cases that have a special status in the research field. Regarding the latter category, scholars – particularly those doing comparative historical analysis – often argue that some cases merit more interest than others. These can be termed “paradigm cases” (Gerring 2001:219), that is, cases any theory would need to make sense of because they are understood to be substantively important for our basic understanding of a phenomenon or “because of a past or current major role in domestic or international politics” (Mahoney and Goertz 2006:242).¹¹ France and Russia qualify as such cases in studies of revolutions, Sweden in studies of welfare states, and Germany and the Soviet Union in studies of totalitarian regimes. With respect to Downing’s theory, Prussia and France would be likely candidates for paradigm cases of what he terms military-bureaucratic absolutism whereas England would serve as a paradigm case for the preservation of constitutionalism.

Although the theory-testing purposes of case selection are most important when the cross-case fit is good, we also recommend that any deviant cases are selected for in-depth study. This goes against the grain of much of the nested analysis literature, where deviant cases are largely seen as irrelevant insofar as the cross-case relationship can be corroborated (e.g., Lieberman 2005; but see also Rohlfing 2008; Schneider and Rohlfing 2013). However, for qualitative methodologists and for scholars in the research tradition of comparative historical analysis, even one deviant case might propel one to reformulate the theory (George and Bennett 2005; Mahoney and Goertz 2006:230-31).

Taking deviant cases seriously serves two purposes. First, scrutinizing deviant cases might reveal omitted explanatory factors that could be included in the explanatory typology, thereby moving the deviant case into a predicted type of an augmented typology (Schneider and Rohlfing

2013; Seawright and Gerring 2008:302), or – less satisfactorily – making it possible to formulate an *ad hoc* explanation for the contingency that has occurred in this particular instance.¹² Second, detailed study of a deviant case could serve the purpose of analyzing if the case actually belongs in the type that makes it deviant or whether it has been misclassified (George and Bennett 2005:252). Returning to the Downing example, our recommendations entail that a case study of Poland is also carried out. Downing’s (1992) own in-depth analysis of Poland shows that the actors, in this case the powerful Polish nobility (the *Szlachta*), were capable of ignoring the structural constraints on which his theory is based, something that can be categorized as an *ad hoc* explanation of the deviant case.

Finally, suppose the explanatory typology cannot be corroborated based on the cross-case distribution (as indicated by a low CR vis-à-vis the random CR). In this case, it makes little sense to select typical cases or deviant cases. In the absence of a resilient cross-case relationship, we know neither what a typical nor what a deviant case is. We can of course identify such cases vis-à-vis the predictions. But as these predictions cannot be corroborated, we are in reality groping in the dark (Rohlfing 2008). In this situation, we recommend researchers to focus on cases that are undisputedly instances of the outcomes of interest (see Beach and Rohlfing 2015).¹³ Furthermore, among the cases with different outcomes, one should focus on those resembling each other as much as possible with regard to variables emphasized by theories previously used to account for the diverse outcomes.¹⁴ These cases provide inferential leverage because they represent full variation in the relevant outcomes and tentative control for competing propositions (see Slater and Ziblatt 2013; Nielsen 2014). A closer look at such “cases whose variation simply cannot be accounted for by extant hypotheses” (Slater and Ziblatt 2013:1313) provides a promising lens through which to identify (combinations of) explanatory factors that can be used for theory building and subsequently examined at the cross-case level.

Typological theory in comparative historical analysis

A number of different bodies of research could be revisited to demonstrate the purchase of our procedures (see, e.g., Elman 2005). But there is an added value of revisiting the research field of comparative historical analysis. Explanatory typologies arguably have more to offer for research agendas that take a “middle position” in what Rohlfing (2012:128) has called the “trade-off between unit-related comparability and generalizability.” This is exactly the aim of comparative historical analysis. Scholars working within this genre regularly use process tracing to uncover causal sequences and comparisons to investigate the generalizability of expected patterns while remaining “respectful of the historical context and details of each case” (Goldstone 2003:51). We show how our procedures make comparative historical analysis more amenable to empirical testing without compromising these strengths.

More particularly, comparative historical analysis has traditionally aimed to establish configurations of stable sets of structural variables and clear-cut outcomes (Mahoney 2003a). The structuralist or path dependency models, which these studies tend to be based on, facilitate the use of relatively simple explanatory typologies.¹⁵ However, at the same time, these works have been deeply attentive to the case level. Indeed, within-case analysis of causal mechanisms is seen as a *must* within this research tradition. More generally, we note that scholars working within comparative historical sociology have been pulled in two opposite directions: towards cross-case comparisons and towards in-depth analysis of mechanisms within cases. The tension between cross-case and within-case considerations identified in the methodological work on explanatory typologies thus also characterizes comparative historical analysis (Mahoney 2004).

Comparative historians have normally risen to the challenge of dealing with the tension between cross-case and within-case considerations by comparing all relevant cases (within certain

scope conditions) on the macro-level using explanatory typologies and by subsequently carrying out process tracing in each and every of these cases. Many scholars working within comparative historical analysis have, on this basis, combined cross-case and within-case analysis to formulate some forceful and impressive theoretical insights. However, critics have, correctly, pointed out that the aforementioned procedures rule out actually testing these explanatory claims. The problem is that the theory is developed to fit the cases, based on an iterative approach where the theory is changed if it cannot be corroborated on the case level (see Mahoney 2004:95). A genuine test presupposes that empirical variation not used to formulate the theory is enlisted. The only appropriate way to test a theory formulated based on a fixed set of cases is therefore to analyze either new aspects of these cases or to enlist new cases (see, e.g., King, Keohane, and Verba 1994).¹⁶ Otherwise, we are not testing theories but merely describing what happens in a set of cases we know in abstract language (substituting variables for proper names). According to Coppedge (2012: Chapter 5), this means that, willy-nilly, most comparative historical analyses are only theory formulating, not theory testing.

The more particular problem here is that comparative historical analyses often do not seek to generalize beyond the cases they examine empirically (Skocpol 1979:287-90; Skocpol 1984:376; Mahoney and Rueschemeyer 2003:7-8). For instance, Skocpol (1979:135) formulates two scope conditions of her theory of social revolutions, viz. i) a “modernizing agrarian bureaucracy” and ii) that the state in question is “incorporated into international fields dominated by more economically modern nations abroad.” Similarly, Tilly (1992:127-30) states that his explanation only pertains to Europe, Ertman (1997:4) that his explanation is only valid within Latin Christendom, Downing (1992:10) that his analysis only applies to countries characterized by “medieval constitutionalism,” and Luebbert (1991:259-60) that the scope condition of his theory is a particular level of societal modernization only to be found in the Western world in the interwar period that he examines.¹⁷

Seemingly, the findings of these analyses therefore cannot be generalized beyond the cases included in the original empirical analyses – on the basis of which the theory was formulated. However, it is not clear that the theoretical insights produced by the scholars listed above need to be restricted to a particular set of scope conditions (Coppedge 2012: Chapter 5). Several scholars have discussed the explanations of particularly Moore ([1966] 1991) and Skocpol (1979) in wider settings (e.g., Stephens 1989; Glynn and Ichino 2014; Goertz and Mahoney 2005; Slater and Ziblatt 2013), a procedure we take as an attempt to investigate the transferability of these theories beyond the original scope conditions and, thereby, to carry out more demanding empirical testing.¹⁸

The guidelines described in this paper can be seen as an attempt to embrace this point as we show how the transferability of typological theories originally designed for a smaller subset of phenomena can be assessed using the measure of cross-case fit we propose below. This can be seen as a way of tackling the problem that comparative historians often select hypotheses to fit the cases and sometimes, probably inadvertently, cases to fit their hypotheses (see Coppedge 2012:148-49). Next, we show how, confronted with the consequent increase in the N, a systematical case selection can direct us to a set of particularly interesting cases on which to test new within-case aspects of the theories developed on the basis of cross-case similarities and differences. This makes it possible for researchers to select a more limited number of cases for within-case analysis, thereby avoiding the often untenable ideal that all cases included into the explanatory typology must be analyzed in depth.

Birth of the Leviathan

Our first example is the one also used by Mahoney (2003a:345-46), namely Ertman's (1997) attempt to explain 18th century outcomes with regard to the regime form (absolutism or constitutionalism) and the character of the state infrastructure (patrimonial or bureaucratic) within

Western Christendom. Ertman explicitly illustrates his dependent variable as a fourfold typology exhausting these two dimensions and, likewise, dichotomizes his two independent explanatory factors, viz., the onset of sustained geopolitical competition (pre-1450 or post-1450) and the type of local government during the first period of state building (participatory or administrative). The explanatory typology can be illustrated as follows.

[Table 2 about here]

Ertman's initial theoretical expectations imply a perfect empirical clustering along the diagonal of the typology. This is so because the presence of the attribute of onset pre-1450 explains the presence of the adjective "patrimonial" whereas the presence of the administrative kind of local government explains the presence of the noun "absolutism" – in a completely symmetrical way. It turns out that only nine of the 13 cases are in fact classified in accordance with these hypotheses, which finds expression in a CR of .69 (9/13). A random distribution would produce a CR of .25, meaning that the performance of Ertman's initial explanation is significantly better ($p=.001$).

However, Ertman is not satisfied with this initial explanation, which, *inter alia*, does not explain the paradigm case of Great Britain. To account for the four misfits, he introduces a particular species of parliaments as an independent causal factor. Ertman's analytical point is that the strong two-chamber parliaments, which were a product of the participatory type of local government in the early phases of state building, could break the causal path created by the other structural variable – the onset of geopolitical competition – whereas the weaker tricursal estates produced by an administrative type of local government could not do so. Hence, the variable that determines the political regime form has systematical knock-on effects on the character of the state infrastructure. Indeed, Ertman (1997:31-32) proposes that the two-chamber parliaments would

(successfully) move their country from patrimonialism to bureaucracy in the case of an early onset of geopolitical competition and, *vice versa*, from bureaucracy to patrimonialism in the case of a late onset of geopolitical competition. Substituting proper names for variables, Ertman argues that the strong English parliament countered the patrimonial tendencies brought about by the (pre-1450) involvement in the Hundred Years' War, in effect moving the English/British case into the bureaucratic category on the dependent variable of the character of the state infrastructure. In Poland and Hungary, on the contrary, the equally strong parliaments resisted the bureaucratic impulse that followed from the late (post-1450) onset of geopolitical competition, in turn, moving the two countries into the patrimonial category on the same dependent variable (see Table 3).

[Table 3 about here]

Accounting for the independent effect of two-chamber parliaments, the CR increases to .85 (11/13). Compared to the random distribution of .25, this is obviously a good fit ($p=.000$), and it is fair to say that Ertman's two-step explanation stands out for both its explanatory purchase and its elegance. The replications furthermore show that Ertman integrates temporal effects into his explanatory typology (Bennett and Elman 2006:467; cf. fn. 5 above). In Tables 2 and 3, this is done in two ways. First, one of the two explanatory attributes is the timing of geopolitical competition. This shows how the effects of timing can be made an inherent part of any explanatory typology. Second, the point about the independent effect of a two-chamber parliament makes up a dynamic element. We have captured this by showing how Ertman's initial expectations can be altered through a conditional relationship. As George and Bennet (2005:245) emphasize, typologies are ideal for capturing such interactions.

Regarding the case selection, Ertman (1997) follows the standard procedure of comparative historical sociology, namely, to analyze all of the cases included in the explanatory typology in depth. With respect to the validity of the theory testing, this procedure is ideal. However, it is also extremely exacting. Based on our guidelines, Ertman could have chosen to select one case from each of the four configurations shaded in Table 3 and the two deviant cases of Denmark and Sweden. Where more than one case is situated in the predicted configurations, either a random selection or more pragmatic considerations could be marshaled to identify the cases. This would have effectively halved the number of case studies without weakening the within-case test very much. Notice, further, that the findings of Ertman's two deviant case studies correspond to one of the uses we highlighted above. In each of these cases, Ertman (1997) has to introduce contingent elements into his explanation based on diffusion (Denmark) or actors' choices (Sweden).

Assessing the transferability of typological theories

In the online appendix, we have replicated two additional comparative historical analyses, namely Luebbert (1991) and the extension of Rokkan by Aarebrot and Berglund (1995). Here, we proceed by considering another study by Ertman (1998), which is apt for demonstrating the possibilities for assessing the transferability of typological theories and, hence, for making comparative historical analysis more amenable to genuine empirical testing. Ertman uses a critical examination of a number of prior works on democratization and democratic stability to present the broad outlines of an explanation of democratic stability and breakdown in interwar Europe. He identifies two crucial factors: the character of the interwar civil society and whether the party system was the pivotal point of the body politics before World War I. Regarding the former explanatory factor, his point of departure is that the civil society has, on the one hand, been connected to democratic stability but

that, on the other hand, scholars have argued that civil society have sometimes facilitated the mobilization of undemocratic forces.

However, following Berman (1997), Ertman argues that these “perverse” effects of a vibrant civil society are conditional on a low level of political institutionalization. He observes that in interwar Europe, civil society only safeguarded democracy where a competitive party system – based on the parliamentary principle – existed prior to 1914 (Ertman 1998:499-500). The Scandinavian countries, Benelux, Great Britain, France, and Switzerland are all instances of this combination (see Table 4). In contrast, as Italy and Germany shows, in spite of a vibrant civil society, democracy broke down where no such competitive party system existed prior to World War I. Likewise, democratic breakdown occurred where political life did revolve around the party system prior to World War I but where no vibrant civil society was found. Portugal and Spain exemplify this combination, the point being that patron-client structures eroded democracy from within. Finally, Ertman uses Russia as an example of an instance in which neither of the two factors were present and where democracy therefore came to naught.

[Table 4 about here]

Civil society and a legacy of party-centered politics can thus be seen as individually necessary and jointly sufficient for democratic stability. If we accept Ertman’s own scoring of the cases and his scope conditions (more on this below), the CR of the typological theory is an impeccable 1.0 (13/13) – to be compared with a random CR of .5 (4/8). Suppose, first, that both the sample and the scoring of the cases are accepted. In this scenario, case selection is relatively straightforward and mirrors the Downing example discussed above. At least three typical cases should be selected for process-tracing, i.e., one from each of the predicted configurations except the overdetermined one

characterized by the combination of weak civil society and no previous experience with party centered politics.

Suppose, next, that Ertman's scope conditions are challenged. Ertman excludes all of East-Central Europe from his analysis based on the argument that these were mostly new states created by the breakdown of empires – which makes it meaningless to examine whether political life prior to World War I revolved around a competitive party system (Ertman 1998:476). For analogous reasons, Ertman also excludes relatively Western countries such as Austria, Czechoslovakia, Finland, and Ireland. However, some of the excluded countries – Czechoslovakia, Finland, and Ireland – were instances of democratic survival during the interwar period. These cases indicate that a competitive party system prior to World War I was not a necessary condition after all. Furthermore, if we introduce a status as a democracy after World War I as the sole – more general – scope condition for the theory Uruguay represents a democratic breakdown despite its pre-World War I experiences with party-centered politics and a relatively strong civil society. In this case, Russia should be excluded while 19 other cases are added. Investigating the transferability of Ertman's theory to interwar democracies *tout court* in this way, we arrive at the case distribution illustrated in Table 5.

[Table 5 about here]

This augmented version of Ertman's explanatory typology produces a CR of .87 (27/31). Statistically, the distribution is still highly significant – with a p-level of .000 – vis-à-vis the random CR of .5. Nonetheless, we now have four deviant cases. With the new scope condition, the three typical cases (or more than three if resources permit) chosen for within-case analysis must – based on our guidelines – be supplemented with these four deviant cases.

Even a cursory review of the literature on some of these cases show that such a study might well justify slightly altering the definition of the party competition variable. Ireland thus had experience with competitive party systems prior to World War I, even if the country was not independent but a constituent unit in the British political system. Irish representatives sat in Westminster where the parliamentary principle was sacrosanct but where suffrage was still relatively limited. In addition, after 1906, Finland had a structured party system, based on universal suffrage, with two poles, a social-democratic and a bourgeois one (Alapero and Allardt 1978:123). The parliamentary principle did not operate as the Russian Tsar designated the government (Karvonen 2000:131), and as such, Finland seems to be deviant for Ertman's explanatory model. However, in-depth scrutiny might well reveal that the legacies Ertman associates with pre-World War I political competition were also in operation here. According to Alapero and Allardt (1978:126), the pre-1914 party system contributed "to the viability of the parliamentary political system established after the civil war" by creating a "clear institutional continuity." An in-depth analysis of the Finnish and Irish cases might thus plausibly either serve as an example of within-case analysis revealing that the two cases have been misclassified or as an example that a slight reformulation of the theoretical model (where structured political competition in a more general sense is substituted for the parliamentary principle) serve to move two hitherto deviant cases into a predicted configuration.

Moore and Skocpol

Two additional examples further serve to illustrate both our procedures and their relevance for considering the transferability of typological theories. As mentioned above, several scholars have sought to examine whether the explanations of Moore ([1966] 1991) and Skocpol (1979) find

support beyond the original scope conditions used to formulate these theories, a procedure we take as an attempt to investigate transferability.¹⁹

In Tables 6 and 7, we have used Goertz and Mahoney's (2005) and Stephens' (1989) reappraisals of Skocpol's and Moore's respective analyses. As in the replications above, we have shaded the expected types based on the theories of Skocpol and Moore.

[Table 6 about here]

Beginning with Skocpol, Table 6 lends strong support to her theory even when additional cases are included. In fact, the CR is no less than .94 (15/16), which is to be compared with a score of .5 for a random distribution, meaning that the better ability to predict the patterns is highly statistically significant (at the .001-level). The theory testing case selection should accordingly be invoked. Analogous to the Downing replication, at least one case from types 1, 6, and 7 should be analyzed in-depth. Also, the one deviant case of Poland-Lithuania (in type 5) should be subjugated to within-case analysis.

[Table 7 about here]

Turning to Moore, the picture changes dramatically, at least if we accept Stephens' (1989) attempt to test the theory.²⁰ As illustrated in Table 7, only three out of the 13 countries included by Stephens (1989) behave as expected based on Moore's theory, making for a CR of .23. To be sure, this should be compared to a meagre score for a random distribution: .13 (2/16). Nonetheless, this version of Moore's analysis with additional cases is not statistically significant. With a p-value of .23, it is far from the conventionally accepted thresholds of .05 or .1. Our measures thus corroborate

the widespread recognition that Moore only got some part of his analysis right (Femia 1972; Skocpol 1973; Stephens 1989; Mahoney 2003b) whereas Skocpol's theory is more resilient (Mahoney 2003a). But the Moore replication also goes to show why it is vital to assess the explanatory power in light of the random fit as the CR is in fact substantially higher than that produced by chance even in Moore's case.

With respect to case selection in the Moore example, we invoke the theory-building guidelines. Cases which are clear-cut instances of the outcomes of interest should be subjugated for in-depth analysis in order to build a new theory that can account for these outcomes. Britain would be one such case with respect to democracy whereas Germany would serve as an example of an undisputed member of the set of autocratic regimes. These two cases can furthermore be seen as paradigm cases, which any new theory would need to make sense of.

Conclusions

The use of explanatory typologies has a strong standing within social science. However, the recent discussions of typological theory pay little or no regard to the general explanatory power as determined by the empirical distribution of cases, and they do not offer a genuine integration of cross-case examination and the selection of cases for within-case process-tracing. In this paper, we have addressed both of these problems. We have proposed a two-step test of cross-case explanatory power. By calculating the coefficient of reproducibility, it is possible to assess the extent to which the cases clump in accordance with the typological theory in the theoretically infused property space – both in absolute terms and vis-à-vis the fit that would be the result of a random distribution of cases.

This two-step procedure enables researchers to appreciate both the absolute explanatory power of rival theories and their respective fits based on the scope of the predicted

outcomes. While we maintain that this procedure is illuminating, we are also keen to add that researchers should not rely unconsciously on the consequent p-levels. The statistical test only reveals a crisp distinction between whether chance could have brought about the pattern or not. But we are very assertive when it comes to defending the logic on which the significance test is premised: It is of the essence to interpret the proportion of correctly predicted cases against the ambitions of the theory. Next, we have argued that a self-conscious case selection is only possible after having taken stock of the cross-case explanatory power of the typological theory. More particularly, when cross-case analysis supports the theory, researchers should carry out process-tracing in cases that allow for a test of the causal mechanisms associated with each of the sufficient conditions, or combination of conditions, for all the outcomes of interest. These in-depth studies of typical cases should be supplemented with within-case inspection of any deviant cases. When cross-case analysis does not corroborate the theory, scholars can use process tracing to facilitate theory building by focusing on cases that are clear-cut instances of the outcomes of interest, in combination with cases with different outcomes but scoring fairly similar on conditions emphasized by extant explanations.

We thus come down in favor of a multi-methods approach where cross-case analysis and within-case analysis are used in a fashion that both allows for testing the macro-level explanatory power of theories and for assessing the underlying causal mechanisms of the theory – and possibly for reformulating the theories based on the subsequent analysis of causal processes. In the literature, typological theory is normally framed in terms of just such an integration of comparative and within-case analysis. However, prior methodological works have not provided clear criteria for how to combine these two elements systematically.

Our procedures furthermore provide assistance in solving two problems of comparative historical analysis. First, our template for cross-case analysis pave the way for

examining the transferability of theories beyond the scope conditions used to formulate them in the first place. In this way, scholars can subject theories to more genuine empirical assessments and, thereby, counter the criticism that comparative historical analysis is only a theory-building rather than a theory-testing exercise. Second, our case selection procedures help scholars avoid the untenable demand that all cases entering comparative historical analysis be analyzed in depth. These guidelines are of course especially important when assessing theories beyond their original scope conditions as the very attempt to investigate transferability increases the number of cases.

References

- Aarebrot, F. and S. Berglund. 1995. "Statehood, Secularization, Cooptation: Explaining Democratic Survival in Inter-War Europe – Stein Rokkan's Conceptual Map Revisited." *Historical Social Research* 20(2):210-25.
- Ahlquist, J. and C. Breunig. 2012. "Model-Based Clustering and Typologies in the Social Sciences." *Political Analysis* 20(1):92-112.
- Alapero, R. and E. Allardt. 1978. "The Lapua Movement: The Threat of Rightist Takeover in Finland, 1930-32." Pp. 122-41 in *The Breakdown of Democratic Regimes: Europe*, edited by J. J. Linz and A. Stepan. Baltimore and London: The Johns Hopkins University Press.
- Anderson, P. 1974. *Lineages of the Absolutist State*. London: Verso.
- Bailey, K. D. 1973. "Monothetic and Polythetic Typologies and Their Relation to Conceptualization, Measurement, and Scaling." *American Sociological Review* 38(1):18-32.
- Bailey, K. D. 1994. *Typologies and Taxonomies*. Thousand Oaks.
- Beach, D. and I. Rohlfing. 2015. *Combining Set Theory and Process-tracing in Multi-method Research: Setting the Parameters of the Debate*. Manuscript.
- Bennett, A. and C. Elman. 2006. "Qualitative Research: Recent Developments in Case Study Methods." *Annual Review of Political Science* 9:455-76.
- Berman, S. 1997. "Civil Society and the Collapse of the Weimar Republic." *World Politics* 49(3):401-29.
- Capoccia, G. and D. Ziblatt. 2010. "The Historical Turn in Democratization Studies: A New Research Agenda for Europe and Beyond." *Comparative Political Studies* 43(8/9):931-68.
- Collier, D., J. LaPorte, and J. Seawright. 2012. "Putting Typologies to Work." *Political Research Quarterly* 65(2):217-32.

- Collier, D. 2014. "Comment: QCA Should Set Aside the Algorithms." *Sociological Methodology* 44(1):122-126.
- Coppedge, M. 2012. *Democratization and Research Methods*. New York: Cambridge University Press.
- Downing, B. 1992. *The Military Revolution and Political Change: Origins of Democracy and Autocracy in Early Modern Europe*. Princeton: Princeton University Press.
- Elman, C. 2005. "Explanatory Typologies in Qualitative Studies of International Politics." *International Organization* 59(2):293-326.
- Ertman, T. 1997. *Birth of the Leviathan. Building States and Regimes in Medieval and Early Modern Europe*. New York: Cambridge University Press.
- Ertman, T. 1998. "Democracy and Dictatorship in Interwar Europe Revisited." *World Politics* 50(3):475-505.
- Fearon, J. and D. Laitin. 2008. "Integrating Qualitative and Quantitative Methods." Pp. 756-76 in *The Oxford Handbook of Political Methodology*, edited by J. M. Box-Steffensmeier, H. E. Brady, and D. Collier. Oxford: Oxford University Press.
- Femia, J. 1972. "Barrington Moore and the Preconditions for Democracy." *British Journal of Political Science* 2(1):21-46.
- George, A. and A. Bennett. 2005. *Case Study and Theory Development*. Cambridge: MIT University Press.
- Gerring, J. 2001. *Social Science Methodology*. New York: Cambridge University Press.
- Gibbons, J. and S. Chakraborti. 2010. *Nonparametric Statistical Inference*. Fifth Edition. Boca Raton: CRC Press.
- Glynn, A. and N. Ichino. 2014. "Increasing Inferential Leverage in the Comparative Method: Placebo Tests in Small-*n* Research." *Sociological Methods and Research*.

- Goertz, G. and J. Mahoney. 2005. "Two-level Theories and Fuzzy-set Analysis." *Sociological Methods and Research* 33(4):497-538.
- Goertz, G. 2008. "Choosing Cases for Case Studies: A Qualitative Logic." *Newsletter of the APSA Section on Qualitative and Multi-Method Research* 6(2):11-14.
- Goldstone, J. 2003. "Comparative Historical Analysis and Knowledge Accumulation in the Study of Revolutions." Pp.41-90 in *Comparative Historical Analysis in the Social Sciences*, edited by J. Mahoney and D. Rueschemeyer. New York: Cambridge University Press.
- Karvonen, L. 2000. "Finland: From conflict to compromise." Pp. 129-156 in *Conditions of democracy in Europe, 1919-1939*, edited by D. Berg-Schlosser and J. Mitchell. Basingstoke: Macmillan.
- King, G., R. Keohane, and S. Verba. 1994. *Designing Social Inquiry: Scientific Inference in Scientific Research*. Princeton: Princeton University Press.
- Lazarsfeld, P. F. and A. H. Barton. 1951. "Qualitative Measurement in the Social Sciences: Classification, Typologies, and Indices." Pp. 155-192 in *The Policy Sciences*, edited by D. Lerner and H. D. Lasswell. Stanford: Stanford University Press.
- Lieberman, E. S. 2005. "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99(3):435-42.
- Luebbert, G. M. 1991. *Liberalism, Fascism, or Social Democracy. Social Classes and the Political Origins of Regimes in Interwar Europe*. Oxford: Oxford University Press.
- Mahoney, J. 2003a. "Strategies of Causal Assessment in Comparative Historical Analysis." Pp. 337-72 in *Comparative Historical Analysis in the Social Sciences*, edited by J. Mahoney and D. Rueschemeyer. New York: Cambridge University Press.
- Mahoney, J. 2003b. "Knowledge accumulation in comparative historical research: The case of democracy and authoritarianism." Pp. 131-174 in *Comparative Historical Analysis in the*

- Social Sciences*, edited by J. Mahoney, and D. Rueschemeyer. New York: Cambridge University Press.
- Mahoney, J. 2004. "Comparative-Historical Methodology." *Annual Review of Sociology* 30:81-101.
- Mahoney, J. 2010. *Colonialism and Postcolonial Development: Spanish America in Comparative Perspective*. New York: Cambridge University Press.
- Mahoney, J. and D. Rueschemeyer. 2003. "Comparative Historical Analysis: Achievements and Agendas." Pp. 3-40 in *Comparative Historical Analysis in the Social Sciences*, edited by J. Mahoney and D. Rueschemeyer. New York: Cambridge University Press.
- Mahoney, J. and G. Goertz. 2004. "The Possibility Principle: Choosing Negative Cases in Comparative Research." *American Political Science Review* 98(4):653-69.
- Mahoney, J. and G. Goertz. 2006. "A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research." *Political Analysis* 14(3):227-49.
- Mikkelsen, K. 2015. *Negative Case Selection: Justifications and Consequences for QCA-based MMR*. Manuscript.
- Moore Jr., B. [1966] 1991. *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World*. London: Penguin Books.
- Nielsen, R. 2014. "Case Selection via Matching." *Sociological Methods and Research*. doi: 10.1177/0049124114547054
- Ragin, C. 2000. *Fuzzy-Set Social Science*. Princeton: Princeton University Press.
- Ragin, C. 2008. *Redesigning Social Inquiry*. Princeton: Princeton University Press.
- Rohlfing, I. 2008. "What You See and What You Get: Pitfalls and Principles of Case Selection of Nested Analysis in Comparative Research." *Comparative Political Studies* 41(1):1492-1514.
- Rohlfing, I. 2012. *Case Studies and Causal Inference: An Integrative Framework*. Basingstoke: Palgrave Macmillan.

- Rokkan, S. 1970. *Citizens, Elections, Parties: Approaches to the Comparative Study of the Processes of Development*. Oslo: Universitetsforlaget.
- Seawright, J. and J. Gerring. 2008. "Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options." *Political Research Quarterly* 61(2):294-308.
- Schneider, C. and I. Rohlfing. 2013. "Combining QCA and Process-Tracing in Set-theoretic Multi-method Research." *Sociological Methods and Research* 42(4):559-97.
- Skocpol, T. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. New York: Cambridge University Press.
- Skocpol, T. 1984. "Emerging Agendas and Recurrent Strategies in Historical Sociology. Pp. 356-91 in *Vision and Method in Historical Sociology*, edited by T. Skocpol. New York: Cambridge University Press.
- Slater, D. and D. Ziblatt. 2013. "The Enduring Indispensability of the Controlled Comparison." *Comparative Political Studies* 46(10):1301-27.
- Smith, K. 2002. "Typologies, Taxonomies, and the Benefits of Policy Classification." *Political Studies* 30(3):379-95.
- Stephens, J. 1989. "Democratic Transition and Breakdown in Western Europe, 1870-1939: A Test of the Moore Thesis." *American Journal of Sociology* 94(5):1019-77.
- Tarrow, Sidney (2010). "The Strategy of Paired Comparison: Toward a Theory of Practice." *Comparative Political Studies* 43(2): 230-259.
- Tilly, C. 1984. *Big Structures, Large Processes, Huge Comparisons*. New York: Russell Sage Foundation.
- Tilly, C. 1992. *Coercion, Capital, and European States AD 990-1992*. Oxford: Blackwell.
- Tilly, C. 1997. "Means and Ends of Comparison in Macrosociology." *Comparative Social Research* 16:43-53.

Tables

Table 1: Downing's typology with empirical cases

	High levels of warfare		Low levels of warfare	
	Mobilization of domestic economy necessary	Other sources of finance available	Mobilization of domestic economy necessary	Other sources of finance available
<i>Military-bureaucratic absolutism</i>	<i>Type 1</i> Prussia France	<i>Type 2</i>	<i>Type 3</i>	<i>Type 4</i>
<i>Preservation of constitutionalism</i>	<i>Type 5</i> Poland	<i>Type 6</i> England II (1688-1713) Dutch Republic Sweden	<i>Type 7</i>	<i>Type 8</i> England I (to 1648)

Table 2: Ertman's typology with empirical cases (without the independent effect of Parliaments)

	Onset pre-1450		Onset post-1450	
	Administrative	Participatory	Administrative	Participatory
<i>Patrimonial absolutism</i>	France Spain Portugal Tuscany	Naples Savoy Papal States		
<i>Patrimonial constitutionalism</i>				Poland Hungary
<i>Bureaucratic absolutism</i>			German Territorial States	Denmark
<i>Bureaucratic constitutionalism</i>		Britain		Sweden

Table 3: Ertman's typology with empirical cases (with the independent effect of Parliaments)

	Onset pre-1450		Onset post-1450	
	Administrative	Participatory	Administrative	Participatory
<i>Patrimonial absolutism</i>	France Spain Portugal Tuscany	Naples Savoy Papal States		
<i>Patrimonial constitutionalism</i>				Poland Hungary
<i>Bureaucratic absolutism</i>			German Territorial States	Denmark
<i>Bureaucratic constitutionalism</i>		Britain		Sweden

Table 4: Explanatory typology of Ertman's (1998) account

	+ Party centred politics before WWI		- Party centred politics before WWI	
	+ Civil society	- Civil society	+ Civil society	- Civil society
<i>Democratic survival</i>	Belgium, Denmark, France, the Netherlands, Norway, Switzerland, Great Britain, Sweden			
<i>Democratic breakdown</i>		Portugal, Spain	Italy, Germany	Russia

Table 5: Explanatory typology of Ertman's (1998) account with additional cases

	+ Party centred politics before WWI		- Party centred politics before WWI	
	+ Civil society	- Civil society	+ Civil society	- Civil society
<i>Democratic survival</i>	Australia, Belgium, Canada, Denmark, France, Netherlands, New Zealand, Norway, Switzerland, Sweden, United Kingdom, USA		Finland, Czechoslovakia, Ireland	
<i>Democratic breakdown</i>	Uruguay	Chile, Greece, Portugal, Spain	Argentina, Austria, Italy, Germany	Bulgaria, Estonia, Poland, Latvia, Lithuania, Portugal, Yugoslavia

Table 6: Skocpol's typology with additional cases (based on Goertz and Mahoney 2005)

	State breakdown		Not state breakdown	
	Peasant revolt	Not peasant revolt	Peasant revolt	Not peasant revolt
<i>Revolution</i>	China			
	Russia 1917			
<i>Not revolution</i>	France			
	Poland-Lithuania	England Japan India Portugal Prussia Spain Sweden	Ottoman Empire Austrian Empire Austria-Hungary Dutch Republic	Russia 1905 Germany

Table 7: Moore's typology with additional cases (based on Stephens 1989)

		Strong landed elite		Weak landed elite	
		Significant bourgeoisie weaker than landed elite	Significant bourgeoisie stronger than landed elite	Significant bourgeoisie weaker than landed elite	Significant bourgeoisie stronger than landed elite
<i>Authoritarian regime</i>	Not revolutionary break	Austria Spain	Germany ²¹ Italy		
	Revolutionary break				
<i>Democracy</i>	Not revolutionary break			Sweden Denmark Norway Switzerland	Belgium Netherlands Finland
	Revolutionary break		Britain		France

Endnotes

¹ A couple of clarifications to situate typological theory within this special issue: Typological theory can in many respects be seen as a simpler version of QCA (see Elman 2005: fn. 15; Schneider and Rohlfing 2013: fn. 4). However, there are also some important differences between the two methods. First, although typological theory includes a number of strategies of reduction originally devised by Lazarsfeld (Lazarsfeld and Barton 1951; see also Elman 2005), it does not employ the QCA manipulation tools, including Boolean minimization and the formalized use of logical remainders (i.e., unobserved configurations). Second, as opposed to QCA, explanatory typologies can simultaneously operate with more than one positive outcome.

² Ragin (2008) has discarded this idea later on and replaced it with the concepts of consistency and coverage. Although Ragin has put aside this conception of quasi-sufficiency in the course of developing QCA, it is still useful for our purposes.

³ See our own replication of Ertman's study below.

⁴ At the 0.10-level.

⁵ Some might question the meaningfulness of using significance tests in comparative historical analysis because the case selection is mostly not random (see Coppedge 2012: 147). Indeed, many scholars working within this tradition claim to analyze the entire relevant population. However, such claims need to be based on specific assessments of the transferability rather than simply assumed, and even if we were able to identify all existing, relevant cases, we cannot rule out the possibility that – in the future – other cases might also be relevant.

⁶ After our initial submission of this paper, we have become aware that Coppedge (2012:147) has proposed to use Fisher's exact test to evaluate if the results of a series of comparative historical analyses could have been produced by chance. The results based on the binomial probability tests and Fisher's exact tests, respectively, are unlikely to differ much with respect to the examples we replicate. In that respect, we see Coppedge's suggestion as complementary to the statistical procedure proposed in this paper. That said, the binomial probability test is arguable preferable because it is a goodness-of-fit measure developed to compare an actual distribution with a theoretical distribution; Fisher's exact test is a measure of independence developed to test if the distribution within two independent groups are different. We also note that Coppedge does not consider how the issue of transferability can be linked with the statistical test.

⁷ To illustrate this problem, we have carried out a simple experiment where we have asked three PhD students trained in comparative analysis to – independently of each other – evaluate three of our examples in an informal way in order to

judge which has the better fit. We asked them to evaluate the examples illustrated in Table 1 (Downing) and Table 2 (Ertman I) in this paper and in Table 2 (Luebbert) in the appendix. Of these, the Downing example has a much lower goodness-of-fit based on the binomial probability test than both Ertman I and Luebbert. But in all three trials, Downing was identified as having a better fit than at least one of the other examples based on the reasoning that his theory predicts a higher proportion of cases (as mentioned above, a similar criterion is used by Ertman in his discussion of prior research). This rank order was the case even in the one trial where the PhD student explicitly stated that s/he had judged explanatory power based on both the proportion of correctly classified cases and the proportion of predicted types of all types, the two *desiderata* of the binomial probability test. This underscores that informal assessments carried out by social scientists are unreliable – and that a more systematic procedure is not a trivial contribution.

⁸ Later in the chapter, George and Bennett (2005: 254) refer to “typical cases.” But even here it is not clear that the case selection follows from the cross-case patterns.

⁹ Some would probably raise the question about equifinality at the level of mechanisms here. We follow Schneider and Rohlfing (2013) in arguing that, pending evidence to the contrary, cases situated in the same type are qualitatively identical with respect to the mechanisms linking the conditions to the outcome. For this reason, in-depth study of one case from each of the relevant configurations is in principle enough, even though more cases studies are of course always better.

¹⁰ See Schneider and Rohlfing’s (2013) principle of unique membership, saying that one should select cases only covered by one solution term.

¹¹ If scholars, due to resource limitations, are not able to carry out all the suggested in-depth case studies, an additional criterion is the number of cases falling under the different types. The logic here is that we want to gain as much analytical leverage as possible by scrutinizing high-frequency configurations.

¹² This point is arguably of more general interest as it shows that the division between theory building and theory testing is often not hard and fast. The testing only makes sense inasmuch as the model is correct, meaning that anything that challenges the model (e.g., an omitted variable) affects the validity of any tests.

¹³ This strategy for identifying the relevant range of outcomes is termed typological representativeness by Slater and Ziblatt (2013) (see also Collier et al. 2012).

¹⁴ In crisp-set QCA, such cases would be placed in the same row, indicating a contradictory configuration.

¹⁵ Anderson (1974) and Rokkan (1970) explicitly stressed that producing such typologies was their objective. But even in studies where typologies are not explicitly mentioned, the analyses carried out can to a large extent be understood in

these terms. As Skocpol (1973) and Stephens (1989) have shown, Moore's ([1966] 1991) distinction between three routes to democracy and dictatorship is an example of an explanatory typology. Also, Skocpol's (1979) own analysis of *States and Social Revolutions* can be presented in this way (see Goertz and Mahoney 2005). In addition, the more recent analyses of Luebbert (1991), Downing (1992), and Ertman (1997; 1998) obviously take the guise of explanatory typologies. We replicate most of these studies in this paper or in the associated online appendix.

¹⁶ It follows from this that if we are able to develop independent observable manifestations of a theory for a particular case or set of cases, i.e., new predicted evidence that was not used to build the theory in the first place, then we can test theories without enlisting new cases. But with the kind of macro-structural variables that pervade comparative historical analysis this will often be farfetched – or at least the new implications will not get at the core of the theory. We therefore concentrate on developing procedures that pave the way for enlisting additional cases.

¹⁷ Downing (1992) and Luebbert (1991) each make the case for their scope condition by analyzing cases not characterized by such conditions.

¹⁸ For a systematic strategy to expand populations by relaxing scope conditions, see Rohlfing (2012: Ch. 9).

¹⁹ Moore's situation differs somewhat from the situations enumerated above as he does not emphasize that his theory is only valid within particular scope conditions. Indeed, "[b]y applying these conditions to contemporary India, Moore indicates that he considers his theory to be relevant to the present (and implicitly the future) as well as the past" (Femia 1972:23). However, Moore does emphasize that smaller countries cannot be used to test his theory since "the decisive causes of their politics lie outside their own boundaries" and that "their political problems are not really comparable to those of larger countries" (Moore [1966] 1991: x). As such, the addition of such empirical referents does violate his scope conditions, albeit in a very different way than in the case of the Skocpol replication.

²⁰ Stephens (1989) actually operates with no less than five dimensions. We have only employed three of them as the two additional factors seem to work as intermediate variables rather than independent explanatory factors.

²¹ Stephens (1989) does not clearly say whether the bourgeoisie in Germany was weaker or stronger than the landed elite. Based on Luebbert (1991), we score it as stronger.