# A Developmental Approach to Historical Causal Inference

David Alexander Bateman, Cornell University

Dawn Langan Teele, University of Pennsylvania

Abstract: Two perennial concerns in the field of political development are why institutions endure and why they change. As scholars increasingly study these questions using the potential outcomes framework, we argue that the older-school understanding of historical sequencing – a developmental understanding of causation – should not be forsaken. In our own research on electoral reform, both the disenfranchisement and re-enfranchisement of Black men in America, and the enfranchisement of women in semi-democratic countries, we note that the relationships between factors that are important for reform may change over the course of history. Some factors may stand out as less relevant in the final instance of a reform even though they were crucial in earlier potential moments of change, and even though they ensured the continued salience of the issue. A full conceptualization of reform therefore requires close examination of the plausible counterfactual moments of reform, and a theory of how the sequence of events in those moments affected the stocks of political variables, as well as their flows, in the final instance. At the same time that causal inference has resonated with historically-minded social scientists, historical institutionalism's emphasis on dynamic relationships can enhance methods of causal inference.

Workshop Participants: This essay is for a special issue of Public Choice on causal identification in American Political Development. The final version of the essay is due May 6. So please point out problems, suggest citations (including from your own work!) and make any other recommendations for the final draft.

## INTRODUCTION

While description and interpretation are both key elements of social science research, the goal to which many of us aspire is the ability to make and empirically substantiate causal explanations for complex social phenomena. We want to know what has happened, what it means, what its consequences were or will be; but, perhaps above all, we want to know *why* it's happened. This is, to put it mildly, a tricky business, and debates about how to best go about doing causal inference have been central to social science since its emergence as a distinctive field of study.

With the rise of a more formalized counterfactual approach to causal inference – the potential outcomes framework (Rubin 1974; Holland 1986; Neyman 1923) – there has been a renewed appreciation for, and use of, experimental and so-called quasi-experimental methods across the discipline. For observational social scientists, the advance of design-based inference has been largely salutary: it has provided those researchers interested in establishing causal effects with a powerful approach, if not exactly a template, for how to design their studies to meet the framework's demanding assumptions; it has emboldened more of us to try and make definite claims about causality; and in general, it has the potential to encourage more deliberateness in the conduct of research.[1]

The advent of the potential outcomes framework has helped drive the "historical turn" in comparative politics and fostered a renewed appreciation for history in the study of American politics.[2] But design based inference has yet to produce a key to all mythologies for research into the past. Its adoption, while rapid, has been uneven, both because of persistent epistemological and methodological diversity among historically-oriented scholars but also because the design-based inference approach is more complementary with some types of research questions than others. Historical social scientific questions come in different types, three of which are particularly common in political science and cognate fields and which lend potentially themselves to quantitative concepts

[1] Indeed, an as-yet unrealized potential of the demanding assumptions of this framework might be to bolster the visibility of non-causal and descriptive research as important contributions to knowledge. If the assumptions of the potential outcomes framework cannot be reasonably met through design decisions, that does not mean the subject is not worth studying. A full appreciation of this fact, we hope, might restore careful descriptive and interpretative work to equal status as contributions to knowledge.

[2] On the historical turn see Capoccia and Ziblatt 2010; on the relationship between American political development and comparative politics see Morgan 2016.

of causal identification. The first of these are questions about "historical legacies," where scholars seek to understand contemporary phenomena in light of long term cultural and economic processes or specific historical interventions.[3] The second might be described as "history as a case" type questions, in which the researcher approaches historical episodes as a data point that can be used to analyze more general theories of politics.[4] The third, and primary focus of this paper, are questions about particular "historical events," in which scholars attempt to establish varyingly comprehensive explanations for the timing and nature of major political transformations by reconstructing the short and long term processes that produced them.

Our intuition is that the strictures of design-based approaches do not lend themselves equally well to all three types of questions. The historical legacies branch is perhaps especially compatible with a potential outcomes or design-based inference framework, in part because temporal distance can be plausibly leveraged for purposes of causal identification. The second type of question is less concerned with history as history than with history-as-data, treating the past as a repository of observations that can be leveraged for either increasing statistical power or providing variation across key variables or parameters of interest. Researchers operating in this genre can often use coarser indicators from the past paired with more finely grained and multifaceted data from the contemporary era. Because they are less interested in understanding the distinctive features of a specific historical moment or event, researchers working in this mode can even supplement historical data with newly constructed data that they can tailor to better meet the assumptions of design-based inference.

The third type of historical research, on major historical events, is arguably the most difficult to study in a design-based framework. Questions about why a particular event occurred or process unfolded are by no means empirically intractable, nor will research into them be uninformed by more generalizable theories. But the very nature of the question makes data limitations more intractable than in "history-as-data" type questions, while the dynamic nature of many of the events means that rather than leveraging temporality for purposes of identification we must integrate its particular and recursive dynamics into our explanations. That is, if the event we seek to explain was not a single interruption into political and social life, but one that unfolded dynamically over time,

[3] e.g. AJR 2001; Voigtlander & Voth 2012; Nunn 2008; Acharya, Blackwell, and Sen 2018; Abramson and Carter 2016.

[4] E.g. Eggers and Spirling 2017; Cirone and Van Coppenolle 2018.

then modeling the causal process requires close attention to the possibility for co-determination in the values of important co-variates.

Of course, historically oriented researchers often operate in more than one of these modes in any given project: those who seek to explain a single event rarely present so idiosyncratic a story that there are no generalizable mechanisms that might be usefully substantiated with contemporary data or newly designed experiments, and the importance of the event in question is often justified by making claims about the long term legacies. Still, we suggest that each of these stands as a distinct type of empirical project, and that while they might be usefully combined in a larger research agenda, they pose distinctive obstacles and raise particular questions from the perspective of research design.

Our paper provides an entrée into understanding some of the common challenges that present in applications of causal inference in the third of these types of questions, the study of major historical events. These issues are close to our hearts, as they are ones that we confronted in our own investigations into why the right to vote was extended to African American men during the nineteenth century, why different countries adopted women's suffrage at different junctures, and why Congress failed to pass robust voting rights legislation in the late 19th century (Teele 2018a, 2018b; Bateman 2018, n.d.; Bateman, Katznelson, and Lapinski 2018). As we both discovered, answering these types of questions is often less about establishing the causal effect of any particular variable than about evaluating the relative contribution of different theorized causes over an extended period of time. The resulting research often veers toward being more case-centered rather than theory-centered, i.e., it is focused on providing a fuller accounting of a specific event or process rather than empirically establishing a generalized theory of how different variables or processes relate to each other across cases (Rohlfing 2012).[5] Because we are interested in explaining a specific case or set of cases, we cannot always assume that observing similar phenomena in the future would provide much additional leverage on explaining the past. What is more, in the serendipitous world where some data exists, it was almost certainly collected with different ends in mind, meaning that to the extent we want to make any quantitative claims, a deep dive into the archives, and original data collection, will almost certainly be necessary even though the resulting dataset will likely fall far short of any "ideal" data.

[5] As Thelen (1999: 371) and others have argued, a key difference between historical institutionalism and rational choice institutionalism lies in how hypotheses are formulated. Whereas HI proceeds from an interest in historical empirical puzzles, RCI is more concerned with how institutions deviate from deductively derived theories of politics.

Moreover, major historical events are best thought of as a dynamic process, for which causal effects are difficult to estimate even under the best of circumstances. Fortunately, many of the methodological assumptions of APD and historical institutionalism are sensitive to the problems dynamic processes raise for causal inference. Yet the actual mechanics of doing research on dynamic processes can be elusive. We articulate a way to design research on major historical events in which we advise scholars to construct a complete *timeline of relevant counterfactual nodes* for each observational unit. An observational unit can be a country, a state, a legislature, or even a bureaucracy. The event to be studied can include the transformation of a political regime, the adoption of a new electoral system, the passage of a particular public policy, or an overhaul of an institution. Although it is possible that the very first time the resulting event was proposed or discussed it was implemented, typically major events go through many rounds of contestation before the final adoption.

Constructing a timeline requires pinpointing all of the times when the event was up for discussion, and then honing in on understanding the dynamics of each of those moments. This serves several inferential functions. The first and most important is that it will guide deeper probing into sources to learn about the patterns of conflict and the nature of public and political discussions in the relevant moments. Re-creating the sequence of events as they unfolded on the ground, which is effectively process tracing, provides a possibility of eventually being able to evaluate the relative importance of different factors at different stages. The thick descriptive knowledge that comes from this form of learning aids in the task of generating causal theories whose assumptions and mechanisms are well-suited to the particularities of a specific historical setting. In addition, construction of these timelines are crucial for any comparative endeavor, as understanding whether observational units are temporally analytically equivalent is crucial for making qualitative causal inferences (Falleti 2013: 162). Finally, knowledge of the relevant counterfactual moments can help in the construction of a more rigorous quantitative research design. As Kocher and Montiero (2016) argue, design based inference generally proceeds by arguing that at least some key causal variable of interest was allocated in an as-if random way. Because the plausibility of these claims rests primarily on idiographic, qualitative, knowledge, the intimate understanding of specific counterfactual episodes that emerge from constructing the timeline will help the researcher learn of opportunities for exploiting a design-based framework and to explore whether the assumptions of the statistical model are potentially realized or excluded in studying a particular process.

We begin by situating the problem of understanding major historical events in the recent literature on causal inference. We describe how construction of a timeline of counterfactual nodes can help to generate qualitative and quantitative insights into the historical event. Tuning next to a discussion of our own work, we describe how we gradually came to appreciate the difficulties and tradeoffs of working in an area where the data that is available or that might be recovered is very rarely of a form that we would have collected had we been designing a contemporary study. The contribution that our projects make to the larger study of political development is related in large part to the state of the data when we entered the field. In the conclusion, we outline what might be called a developmental perspective on causality.

## CAUSAL INFERENCE AND HISTORICAL RESEARCH

The first authority on causal relationships was John Stuart Mill. Mill argued that we can understand both the causes of effects and the effects of causes if we are able to isolate either different causes and see which effects follow, or different effects to see by which causes they were preceded. By varying the circumstances in which a phenomenon is observed, we can begin to understand whether a cause always has the same effect, or whether effects always derive from the same causes. In other words, by trying to observe relevant counterfactuals we can make inferences about causal relationships. Scholars working in both the qualitative and quantitative traditions have long invoked Mill's type of counterfactual reasoning, but it wasn't until the late 1990s that formal versions of this logic became current.[6]

Although Mill believed that the human mind was capable of evaluating counterfactuals and adducing causal relationships, later scholars have become more circumspect, insisting that controlled experiments are the only way to know. Despite what the vanguard of the randomista cadre would argue (Green and Gerber 2014), there is a lively debate about whether experiments are the only way of knowing (Teele 2014). Ultimately, most political scientists understand that not all things worth studying can, or should be, put into an experimental framework. Nevertheless, the past decade has

[6] There is a large literature on the use of the so-called methods of similarity and difference that derive from Mill's work. In the method of agreement, we study outcomes that are present in all cases. In the method of difference, we study outcomes that are present in some cases and not and others. Mahoney's (2000: 392) instructive article argues that the method of agreement can help us to eliminate necessary causes while the method of difference can help to eliminate potential sufficient causes. The important thing is to subject our favored hypotheses to tests that are as strenuous as those we apply to alternative hypotheses.

seen a flowering of research that attempts to apply the logic of potential outcomes framework to observational questions (Dunning 2012).

To understand the logic of causal inference, imagine that the world is composed of multiple units (people, villages, policy domains) that can be catalyzed by different causes (assignment to treatment or control) and wherein different effects (outcomes) can be observed. In the potential outcomes framework, each observational unit can be theorized as having multiple potential outcomes depending on the treatment that it receives, but in a single-shot experiment each unit can only be assigned to one treatment group, and hence will only have one realized outcome. The fundamental problem of causal inference is that we can never directly observe a causal effect, because only the outcome, and not both the outcome and potential outcome under a different treatment, can be observed. As a result, a host of assumptions are required to produce unbiased estimates of causal relations. In general, the quantitative identification of a causal effect relies on the assumption that the assignment of observational units to the treatment or control groups is independent of potential outcomes. For those questions that can be ethically answered in an experimental context, the math (if not the logistics) are relatively straightforward: randomization ensures that the assumption of independence holds, in which case the expected value of the unobserved potential outcomes will be equal to the expected value of the observed values, conditional on treatment.

*The Fundamental Problem of Historical Causal Inference*

Historical work throws up several obstacles to causal inference in this framework, but the basic difficulty is easily stated: historical analyses rely on data that is necessarily observational and usually post hoc. As a result, we are unable to rely on one of the most powerful means for causal identification – the deliberate randomization of exposure to treatment. For this reason, scholars have increasingly turned toward what is known as "Design Based Inference" in order to make causal claims about historical processes, that is, an intentional effort by the researcher to address the difficulties of causal analysis through design choices rather than statistical modeling (Imbens 2010, 403; Keele 2015). In practice, this means attempting to reduce heterogeneity within study groups, and to examine implications of the requisite identifying assumptions, especially those related to the assignment of units to different treatment categories (Dunning 2012).

There are several types of research that use historical data: studies that attempt to explain contemporary phenomenon as the legacy of long-term processes; those which use history as a data

point to examine general theories of politics; and those which seek to understand major historical events. As we suggested in the introduction, the third type of historical research, that on major historical events, may be the least well suited to design-based inference. The problems include how researchers conceptualize the "event" and the types of data that they will be able to bring to bear in a causal inference framework. While major events generally have long-term legacies, and their occurrence may provide some insight into generalizable theories of politics, the emphasis in this line of research remains at least as much on understanding the particularities of specific "case"—why it happened in the way and at the time that it did—as with the universe of cases to which a hypothesized theory might apply.

This has important implications for our ability to acquire the data that we might want or to rely on the guidance of theory developed with reference to contemporary phenomena.[7] For instance, it is not usually the case that the limits of historical data can be overcome by the addition of newly *generated* data that better meets the standards required by the research design. And insofar as the particular events or processes we are interested do not have the benefit of an already established and systematically organized infrastructure of data and knowledge—for example, careful empirical work into measuring and describing particular aspects of historical phenomena[8]—our research designs will face hard constraints in the types of data and assumptions that they can realistically hope to leverage.

More fundamentally, however, is the question of how we conceptualize the relevant outcome in this type of analysis. As Paul Pierson (2004 [p?]) and others have stressed, many of the major historical events that we are likely to be interested in are better thought of not as *events* but as *processes* that unfold in historical time, in which the sequence in which variables appear or take on certain values matters, and the relationship between variables at one moment in time will alter their relationship, and their effect on the outcome, at a subsequent moment.

---

[7] This matters because the degree to which "new" cases will occur or "new" data can be created—and not simply compiled—has different implications for our research design strategies. Unless the researcher can make a persuasive case that what they are studying is time- and context-invariant, then designing studies with an eye to producing and collecting genuinely new data based on future events will likely not be an option. More generally, theories derived and substantiated by reference to contemporary work, while a valuable starting point for historical analysis, cannot always be simply applied off-the-shelf to the past, because core assumptions about political behavior and institutional operation might not be valid in this different terrain. As a result, we often lack context relevant theory or empirical guidance about the possible relationships between different variables, making it much more difficult to adequately model the causal process.

[8] See, for example, the discussion of the work on English medieval villages in Carus and Ogilvie (2009).

*Dynamic Causal Processes*

A *dynamic causal process* is one in which the relevant actions or variables that influence an outcome of interest do not occur just once but as part of an unfolding sequence. Dynamic processes pose thorny problems for causal inference. As Blackwell (2012) discusses, examining temporally-unfolding processes as "single shot" phenomena, a common and often unacknowledged practice, will require the researcher to control for variables whose values have evolved over time in relation to other variables or processes of interest. This can amount to conditioning on posttreatment variables, which introduces an important source of bias that will impede the evaluation of causal effects in a quantitative framework. Since sequencing is often key for understanding historical events, statistical analysis of event snapshots will often be hard pressed to meet the basic assumptions of the potential outcomes framework. More generally, whether a study relies primarily on qualitative or quantitative modes of evidence, the arbitrary temporal bounding of dynamic processes is likely to result in the misestimation of any particular causal effect as well as the relative importance of different hypothesized causes (Pierson 2004 [p.?]). More and better data can help us address the problem of dynamic inference, yet, as we discuss below, this can be especially difficult in studying historical events, and inevitably produces tradeoffs between breadth and depth in the research design. Put simply, dynamic causal processes are characterized by the fact that relationships between variables at different moments in the sequence influence their values and causal effects at subsequent moments, complicating our ability to draw static causal estimates for variables of interest.

Consider the issue of electoral reform, or "why was the right to vote expanded" in a particular instance or to a particularly defined group or at a specific historical juncture. A common feature of many of these episodes of democratization is that enfranchised representatives and political elites, in legislatures and elsewhere, debated multiple versions of proposed reforms, and voted on different alternatives and amendments, often over a period of years if not decades. While on its face this might seem to facilitate empirical causal inference by expanding the number of cases, it also generates unappreciated complications by transforming the event in question from a discrete "unit" that occurs more or less all-at-once into an evolving process that unfolds dynamically over time.

A researcher interested in understanding why a reform materialized might, for example, reasonably look at why legislatures and legislators voted for it. But establishing the importance of any particular consideration that legislators might have used in their evaluation of reform proposals

will be hampered by the simple fact that different considerations might matter differently at different moments, with the importance any particular consideration might take at the end of the process often depending on how legislators had responded to other considerations earlier on.[9] Perhaps our researcher hypothesized that legislators' decision to vote for or against the reform was a response to local activism by organizations and social movements who had taken a position on the reform (Bateman forthcoming). An obvious confounding variable might be that legislators were responding primarily to a more diffuse pre-existing public opinion held by already enfranchised constituents, which might also have influenced the likelihood of activist organizations emerging in a district. The hypothesized relationship is represented in the upper panel of the rudimentary path diagram in Figure 1, with $P_d$ indicating district-level public opinion, $A_d$ the level of district-level activism, and $V_i$ the vote cast by legislator $i$ in district $d$.[10] While this is a simplified example, its basic setup has long been a common one in analyses of legislator vote choice: assuming that this model is fully specified—i.e., that we have correctly measured and controlled for all variables that might cause our estimates to be biased—the researcher could claim to have identified a causal effect of local activism, as treatment assignment is now random conditional on the covariate of public opinion.

The historical researcher who discovers that this was not the only time the reform had been voted on, and that it had been the subject of recurring debates for years before final passing, might initially be thrilled by the ability to accumulate more data against which they can test the robustness of their findings and validate their claims about the causal importance of local activism. After all, one of the main difficulties of conducting historical research of particular events is that we often lack—and are unable to produce—data for the cases that we are interested in; expanding the number of cases under study might initially seem a promising avenue for resolving this problem.

But on closer reflection, the discovery that there were previous moments in which reform was debated should in fact complicate our ability to estimate a causal effect for either public opinion or local activism, precisely because it suggests that the event in question was the culmination of a longer and dynamic causal process.

---

[9] The fact that very similar structural situations can, in some cases, drive political actors to reach a *different* conclusion than they did previously also has implications for the use of game theory for understanding historical events, and in particular the application of Markov Perfect Equilibrium techniques which require that, when faced with a similar node in the game tree, actors always make the *same* choice.

[10] This figure is adapted from Blackwell (2012). From a qualitative perspective similar issues have also been explored by Pierson (2004).

We present this schematically in the lower panel of Figure 1. While public opinion might be an important confounder in assessing the importance of local activism on legislators' vote choice, it is also plausible that earlier local activism had an effect on public opinion at a later date. That is, the value of $P_d$ at period $2$ might be a function of $A_d$ at period $1$, even though it itself was a function of $P_d$ at period $1$. Public opinion and local activism, in this case, are not simply connected by a dyadic link, without regard to history, but have developed over the course of an interactive process in which the value of public opinion at one period will be a function of the effectiveness of local activism at an earlier period. This dynamic process will severely compromise our ability to estimate a single causal effect for either activism or public opinion: put simply, if the researcher continues to control for public opinion then they will be introducing a source of posttreatment bias, but if they fail to do so they will be introducing omitted variable bias (Blackwell 2012, 3).[11]

[11] For instance, if local activism makes district-level public opinion more supportive of a reform, then we will underestimate the significance of activism on legislators' voters by including public opinion, but will likely overestimate its significance if we exclude it.

*Figure 1(a): Single-Shot Analysis of Activism on Vote Choice with Public Opinion as Confounding Variable.*

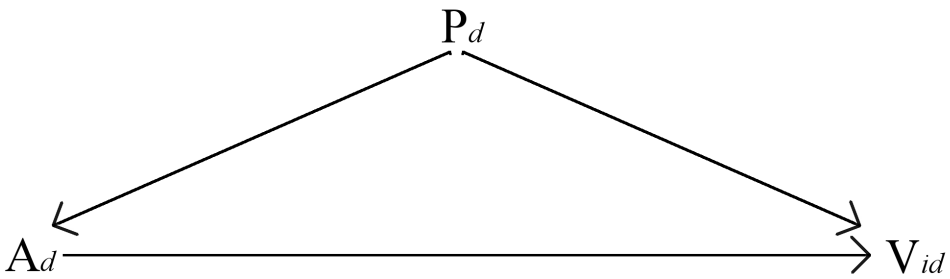
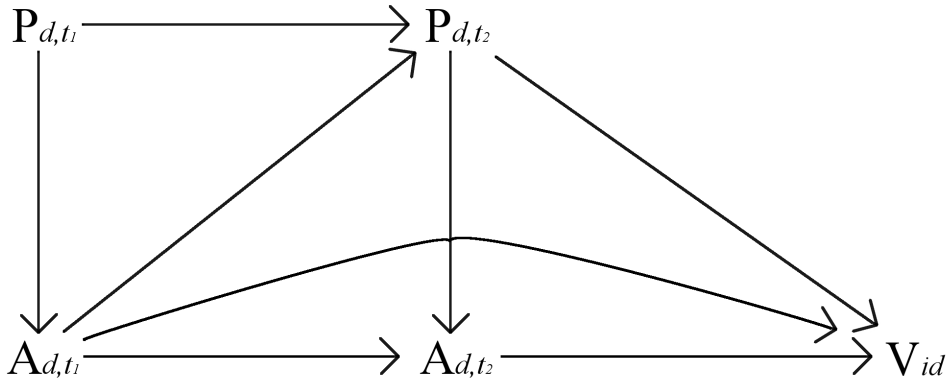



*Figure 1(b): Dynamic Causal Process, with Public Opinion influencing Activism and Vote Choice and Activism at $t_1$ influencing Public Opinion and Vote Choice at $t_2$.*




Adapted from Blackwell (2012)

*Figure 1: Path Diagrams of Confounding Variables*

In one sense, this is a general feature of any process in which an action or treatment occurs not in a singular instance but over time and in relation to other variables of interest. For example, instead of public opinion and local activism we might hypothesize that conditional on ideological orientation, partisan affiliation was most important in shaping legislators' vote choices. But if ideology and party have become gradually sorted over the years-long process of debating this issue, with racial "liberals" becoming Whigs over the issue and racial "conservatives" becoming Democrats, then we are likely to overestimate the importance of party relative to ideology over the course of the process.

Of course, researchers might very well be interested in the respective contribution of causes *only* at the *final* moment of a longer historical process. But this type of intention is rarely specified, and it is not always appreciated that by limiting the question to this specification—in short, by defining the question of interest as a "single shot"—they are also limiting their ability to make more generalizable claims about the importance of different factors to an outcome's occurrence. Much of social inquiry, however, takes the generic form of attempting to evaluate the relative importance of different causes or estimate a generalizable causal effect, and so will be vulnerable to the difficulty this imposes for causal inference. Put differently, understanding why support for a reform went from 45 to 55 percent support in a legislature will be essential for understanding why it passed; but we will have misestimated the relative importance of different causes if we fail to appreciate that some factors that were less important in this final push were more important in getting legislator support up from 1 to 40 percent and placing the issue on the agenda in the first place.

Naturally, the very generalizability of this problem also means that there is nothing distinctly historical about it, and different solutions have been proposed for studying this problem when it occurs in a contemporary setting. But making dynamic analysis central to research design is made more difficult for historical researchers by our inevitable reliance of on observational data in whose production we have had no say. A design-based approach to studying causality, for instance, would require us to take account of the unfolding and dynamic character of the process, but at the outset of a research project we will often lack the necessary tools for doing so, and we lack methodological guidance for how and whether we need to accommodate the possibility that certain relevant variables act as "stocks" of previous rounds of play and not merely as "flows" relevant only at the moment of reform.

To summarize, the fact that there are often multiple moments in which electoral reform was up for debate means that scholars of major historical events can expand the number of observations they use to gain inferential leverage into these complex processes. Yet this discovery will also complicate their ability to estimate a single causal effect for any relevant variables insofar as the respective influence of potentially confounding variables might have changed over time and endogenously in response to earlier events. As we suggest in the next section, creating and then analyzing a timeline of relevant counterfactual moments will help to overcome some of these issues, if only because the relationships between the reform and various independent variables can be evaluated over time.

## THE TIMELINE OF RELEVANT COUNTERFACTUALS

There are two ways that major historical events have been studied quantitatively in the past, both of which tend to think of major historical events as a one-time change. The first is to collect repeated cross-sections for specific observational units (legislatures, countries, organizations) and to measure a host of variables at intervals. In fixed-effects regressions, each observational unit's values in a given cross-section is compared to average values that variable takes in other time periods, the "within" estimator. Thus relative changes in independent variables over time (a flow), rather than the stock of those variables, add to inferences about correlations. Event history studies, on the other hand, give independent weight to the passage of time (where the temporal relationship can be modeled in various ways), but they perform much better when every other observed variable is time-invariant. In other words, the event history analysis treats relevant covariates as a stock. As we have just argued above, both stocks and flows of political phenomenon are important to dynamic processes, and getting a handle on these relationships is key to understanding causality in major historical moments.

To account for these issues, we propose that researchers construct a complete *timeline of relevant counterfactual nodes* for each observational unit, and then analyze the relationships between suspected key independent variables and the outcome (some failed and some successful) at each of these nodes. In other words, the dependent variable has to be re-conceptualized. Our explanandum is not about whether an event did or did not happen in a particular year, but rather why, in a year in which it might plausibly have happened, it did or did not. To do this we must re-create the sequence of events as they unfolded on the ground, in the service of eventually being able to evaluate the relative importance of different factors at different stages. This will often require going to primary source documents, like minutes of national legislatures or "blue books" for US state legislatures, or even to writings of political theorists, to create a literal timeline demarcating all of the years or legislative sessions when a given reform was debated. This research will establish "nodes" in the history of an event in which the reform was plausible, providing information about relevant counterfactuals moments (Simon 2014).

The timeline of relevant counterfactual nodes serves several inferential functions. The first and most important is that it will guide deeper probing into sources to learn about the patterns of conflict and the nature of public and political discussions in the relevant moments. Qualitative understanding of the politics surrounding the nodes of potential reform will help researchers to

generate "causal process observations" – observations about the actual political dynamics in the moments when reforms were debated (Haggard and Kaufman 2012) – with the goal of abstracting away from the particular case at hand to formulate more general hypotheses. If one is interested not only in explaining a single historical event, but how similar events unfolded in other contexts (or cases), the timeline is further crucial for helping to establish which potential comparative cases are temporally analytically equivalent (Falleti 2013: 141). Finally, the timelines of relevant counterfactual nodes serves in statistical inference. When we have a sense of the periods in which an outcome was possible, we can begin collecting data with an eye toward empirically examining the relationship between the outcome and key independent variables in those moments. We discuss each of these issues in turn.

*Qualitative insights from studying counterfactual nodes*

The first inferential function of constructing the timeline of relevant counterfactuals for each observational unit is qualitative. Researchers interested in historical events will quickly discover and come to know the existing research infrastructure on their topic—the amount, type, and quality of the data that has already been produced by long-term communities of scholars. While such infrastructures are inevitably of varying quality, their existence will be invaluable to any historically oriented social scientist, especially those studying dynamic historical processes. In the case of legislative reforms, for example, knowing when bills were debated and voted upon provides a window for looking into newspaper archives, for delving into the personal correspondence of movement leaders and legislators, and also provides a framing for reading parliamentary minutes. These dates also can help guide more selective searches into secondary historical literatures, as historians will often mention these things as asides in projects unrelated to our interests. Pinpointing the nodes of potential reform and gaining substantial familiarly with primary and secondary source materials surrounding these nodes becomes the crucial material on which we draw context-based causal inferences about specific historical events. From this fine-grained knowledge we can then begin to telescope back out to more abstract thinking about historical events. That is, we can begin to create more generalized hypotheses about the way that the reform unfolded across time and space.

It will often be infeasible to generate causal process observations for all nodes on the timeline for all observational units. But consider again our running example of electoral reform: because of the way that reforms generally unfold in a legislative setting, studying the complete

timeline of counterfactual nodes within a specific observational unit—i.e., the occasions on which a specific state debated—will almost always contain both negative and positive cases. This is true even for those units that do not ultimately adopt a reform, because most parliamentary systems require electoral reform to gain support at several different institutional levels before its ultimate adoption.[12] Importantly, too, it is not necessarily the case that the only positive case is temporally the last one. Sometimes particular parties are in favor of reform and are able to secure majorities in certain chambers when they are in power but not when they are out of power.[13] Hence examining the complete timeline of counterfactual nodes within an observational unit is liable to produce insights about both why reforms gained support and why they failed within that case.

In addition, as many comparative scholars have argued, understanding how reforms unfold over time – both in terms of the historical epoch and the sequencing of changes – is crucial for determining whether different cases are analytically similar to one another. It is only within the context of analytical similarly (or even more strongly "temporal unit homogeneity") that qualitative inferences can be made about the underlying causes of the reform or the results thereafter. A key agenda in the institutionalism literature has been related to understanding whether, for example, policy feedback loops or ratcheting effects are related to the temporal space in which change took place, hence for thinking about how things like order, tempo, and xxxx [ ] matter, establishing the timeline of relevant counterfactuals is pre-requisite (Gryzmala-Busse 2010; Falleti 2013).

*Statistical insights from studying relevant counterfactuals*

There are at least two ways in which working through the timeline of relevant counterfactual moments will help researchers engage in statistical inferences about causation. The first is that, as Kocher and Montiero (2016) argue, design based inference generally proceeds by arguing that at

[12] For example, upper chambers have to pass the law, referenda are often required, and presidents or governors might veto. This means that lower chambers can vote yes but the reform may still stall. In many American states, alterations of the suffrage require a constitutional amendment, which often needed to be passed at two successive legislatures and then sent to the public for approval.

[13] The "last" case of voting on the 15th Amendment in New York State, for example, was the effort of the Democratic Party to recede from that state's ratification, in the hope that this would take the ratification effort below the required threshold of three-fourths. In this case one must be especially careful in evaluating what legislators believed that they were accomplishing. If both parties believed that it was effectively too late to stop the amendment, then the votes might have been solely expressive—a way for Republicans who had voted no to make good with their constituents and leader, or Democrats who had voted no to curry favor with the new electorate or signal moderation to voters—without worry of it changing the end result. Again, being able to make this evaluation will rest on highly idiographic knowledge based on deep immersion in the literature and in the case.

least some key causal variable of interest was allocated in an as-if random way. Because the plausibility of these claims rests primarily on idiographic, qualitative, knowledge, the intimate understanding of specific counterfactual episodes that emerge from constructing the timeline will help the researcher learn of opportunities for exploiting a design-based framework and to explore whether the assumptions of the statistical model are potentially realized or excluded in studying a particular process.

Second, identifying potential moments of reform and exploring the relationships between possibly key causal variables will allow for researchers to assess whether the statistical relationship between these events is the same over time, or whether it changes in particular directions. Of course, in the process of generating a timeline of relevant counterfactuals one will be able to identify existing data or opportunities to reduce archival or other qualitative data to quantitative form; ideally, the data that is produced will be of sufficient high quality and granularity that existing statistical causal inference approaches to dynamic causal processes could be used.

This, as we discuss below, will often be unlikely. In that case, researchers might instead track an association between different variables over time, evaluating both statistically and from qualitative and primary sources whether and how their relationship might be dynamically evolving. The cost of doing so is that we will no long be able to generate a single causal estimand. And the need to track a case temporally will greatly limit our degrees of freedom, often making a fully specified model even more infeasible than usual. But estimating a causal effect in a fully identified statistical model in is not the only way we judge cause-and-effect; by tracing the development of a causal process across many different types of data—a form of triangulation, for instance (Rothbauer 2008)—and connecting this to a set of theoretically structured hypotheses, we might be able to make a persuasive case about the relative weight of different factors at different times, even if we cannot say we have identified or estimated a single causal effect.

*Caveats: Data Scarcity and Sparsity*

As we have argued, the construction of timelines of relevant counterfactual moments can provide both qualitative and statistical leverage against the problem of dynamic causation. This is because a central requirement for studying any historical process as a dynamic process is that we are able to measure variables not as unique values at a specified moment in time but in their relevant history. This in turn requires either an already established infrastructure of data and timeline of its

production, or requires the researcher to engage in constructing their own timeline of relevant counterfactuals to identify the sequential moments or nodes where data needs to be collected.

In either case, however, the historical researcher interested in specific historical events and processes will confront the problem that the data that is available or that might be recovered is very rarely of a form that we would have collected had we been designing a contemporary study; as a result, the data often falls short in a number of consequential ways for making causal inferences.

In any case, we will be very lucky or have worked hard to produce even this limited slice of data, and with some exceptions we will very often lack the granularity required to easily study a dynamic historical process as it unfolds.

The exceptions, however, are worth noting, because they highlight some of the tradeoffs that are inherently involved in making causal inferences in historical settings. There are some research areas that have been so effectively covered by cliometricians and historians that there is already a considerable infrastructure of knowledge and data for researchers to draw upon. The decades of work done by the ICPSR, Keith Poole, and others to collect and clean up the voting records of members of Congress, for instance, gives historically-oriented scholars of Congress an extremely important advantage over scholars of state legislative politics for the same period; so too does the fact that Congress, as a body, has always been relatively concerned with preserving and making accessible the documents involved in its daily activities (which can be compared with a state such as Rhode Island, much of the legislative documents for which remain accessibly only in handwritten versions in the state archive).

Generally, the great advantage of historical work is the archive, curated[16] collections of documents, often cotemporaneous with the relevant event or processes, which were at the time very frequently private, confidential, or even just experience-near.[17] This is a huge advantage available to historically-oriented social scientists whose research methods lead them to produce new data or uncover new documents.[18] At its best, the archive allows the researcher to trace a particular causal

[16] The quality of the curation can vary, but if there are archivists—and not just troves on eBay, whose curation has been undertaken by more haphazard processes—then the quality is usually quite high.

[17] By experience-near, we just mean that even if public they were the types of documents that the relevant actors had at the time while they were making their decisions and the types of documents that somehow managed to make it into the archive.

[18] Consider, for example, that despite the fact that we currently have a White House that leaks like a sieve, we do not, at this moment, know nearly as much about the internal deliberations of the Trump presidency as we

process in considerable detail, often at the level of granularity that modeling a dynamic causal process requires.[19] We might not have polls or regularly updated counts of local organizations, but we might instead have the letters and journals of legislators as they put voting rights on the agenda, or, more commonly, see it put on the agenda, and begin to decide what appeals they will make or position they will take. The tradeoff is that we will likely be able to examine only a much more limited but more detailed slice of the total picture: instead of examining all members statistically, we are able to examine only a very few, and are reliant on evidence whose persuasiveness rests ultimately on the strength of the impression it makes or a lengthier and well documented argument about the relevant importance of different factors.[20]

The degree of data scarcity varies across different lines of historical inquiry: we are indeed often able to draw on more reliable and fine-grained measures of the different variables of interest, have the benefit of context-specific theory and empirically substantiated relationships, and often have the additional (and often lacking for contemporary work) benefit of recorded insights into the motivations of key actors. The data might still be not what we would want or would have collected, and much of it might be difficult if not infeasible to reduce to a quantitative form.[21] But if this infrastructure exists, we will be in a much different situation as researchers than if we need to go about creating much of it by ourselves.

Still, in order to study a historical phenomenon of interest as a dynamic process, we will need to use whatever guidance existing research provides to begin mapping out how this process unfolded and to identify the relevant moments that might merit closer study.

---

do about the Nixon administration – and we know more about the Nixon administration than most, given that *everything*—personal, private, public, and criminal—was effectively seized by court order.

[19] Still, the problem that the data was produced for different reasons than our own continues to apply. For example, the ideal evidence a researcher might hope to recover about a political leader's evolving rationale for a decision might be their tortured musings in a private diary. The actual data, if the researcher is *lucky*, will be whatever the leader decided was worth including in a journal that they were carefully preserving in the expectation that it would be preserved for posterity.

[20] Presumably we would want to look at the issue entrepreneurs and those who were most active in debates. Of course, this means that the marginal legislators who were critical for its passage will go unexamined, biasing our causal estimates by overweighting the motivations of the most passionately involved.

[21] Perhaps part of the reason that the study of history continues to be more associated with qualitative approaches is that much of history, for those who experienced it and left traces of their experiences, was qualitative and impressionistic: legislators in the early nineteenth century spoke regularly of public opinion, but with a few exceptions—the gradually emerging industry of the political almanacs, for instance—their access to public opinion was impressionistic, based on conversations with local figures, with other members of the legislature, and emerging from local party meetings.

## DYNAMIC PROCESSES AND DATA SPARSITY IN PRACTICE

In our own research into black suffrage and the antecedents of women's enfranchisement we have both confronted the problems of compiling and measuring scarce data for what we eventually came to see as dynamic processes, albeit in different ways. It is perhaps in the spirit of this paper that we lay bare our own efforts to resolve these issues and the process by which we gradually came to see the events we were studying as dynamic ones requiring distinctive approaches to causal inference.

Working separately, we both began to understand that our first intuitions about how to study the passage of suffrage reforms in a statistical framework – to look for final dates and use an event-history framework – failed to appreciate or adequately model the sequencing of the historical process that led to reform. Instead, we were gradually pushed to the recognition that the fact that these suffrage bills had long histories inside and outside of legislatures which required us to reconceive of these events as dynamically unfolding processes.

Yet because of the very different prior knowledge and infrastructure that existed in each of these research terrains, we confronted very different data landscapes. This had big implications for the allocation of our time in the research process, and for the resulting products of our work. Two vignettes follow outlining the genealogy of our projects and the strategies that we used to overcome particular limitations in the data landscape.

### *Bateman: sparse data and the tradeoffs of research*

Bateman's initial plan for a study on African American voting rights began with a rough timeline of suffrage reform across American history. The basic trajectory that I was interested in is shown in Figure 2, which tracks the growing number of states that disenfranchised African American men, the passage of the Reconstruction Acts and 15th Amendment that removed all explicit racial qualifications for voting rights, and then a smaller subset of southern states gradually re-disenfranchising the vast majority of African American men (and later women). The initial plan was for a quick explanation of the first wave of disenfranchisement that ended in the mid-1830s, a more substantial examination of the political dynamics that undergirded the expansion of the right to vote between 1866 and 1870, and then a conclusion that focused on how the changed legal and institutional environment after the 1870s resulted in different set of coalitions and political strategies for the second wave of disenfranchisement.

That study never happened. Instead, I started digging around in the antebellum era and never quite dug myself all the way out. The reason had much to do with the infrastructure of available knowledge and data, and the complications that arose as I tried to expand on these. What I found was that not only was there relatively little written about the initial period of disenfranchisement, and almost no data on its occurrence beyond a few timelines which turned out to be mostly incorrect, but that the period I was planning on skipping over entirely – from approximately 1835 to 1865, which in Figure 2 is that extended period where nothing is happening – had in fact been the period that produced the most amount of speeches, writing, legislative voting, public voting in referenda, and petitioning on the issue of black suffrage.[22] An issue that exploded onto the political agenda in 1867, and remained at a fever pitch for the next three years at least, had already been a feature of political debate and the subject of intense legislative fights at the state level for at least two decades.

[22] At some point, I came across what I believe remains the only monograph treatment to have systematically examined black suffrage as a political issue in this period, a dissertation whose author died before its completion (Olbrich 1912); the affinity was instant.

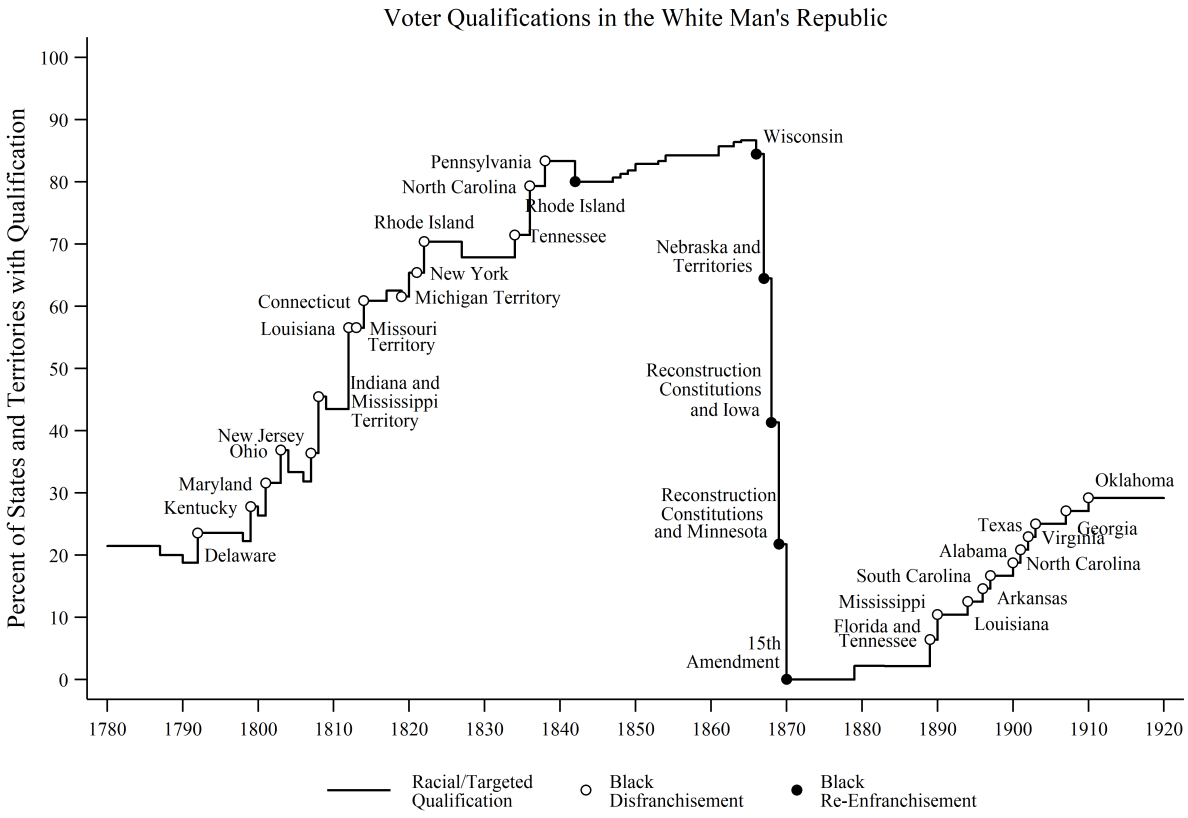



*Figure 2: Timeline of Reform on Racial Voter Qualifications*

Certainly, black suffrage before the Civil War was a different thing than it would become afterwards: enfranchising the relatively small free African American proportion of the population in any given state was not the same as expanding the right to vote to millions of newly freed men in the South. But almost all of the arguments that were being deployed in 1867 for and against black suffrage had already been made; numerous states had already experienced intense political organizing on this issue – as well as murderous backlashes; and northern state legislatures and constitutional conventions saw majorities or near majorities of Republican legislators vote in its favor between 1856 and 1859. It even seems likely that in at least two northern states large majorities of Republican voters had voted in favor of equal voting rights for men by 1859, a fact that was established by the studies focusing on the respective referenda (Field 1982) but which is often glossed over or entirely mischaracterized by much of the secondary literature (see, for example, the treatment in Blight 2018).

It is not that no one knew of this earlier experience. Most accounts of black enfranchisement post-Civil War, for instance, acknowledged the importance of local organizing, and some gestured toward a longer process of contestation (Wang 1997). But historians of the antebellum era had not systematically mapped out how black enfranchisement was put on the political agenda of most northern states, connected to the antislavery cause in rhetoric and public opinion, and anchored as the radical but by no means fringe position of the Republican party well before the Civil War. Lacking this infrastructure, historians of black enfranchisement generally acknowledged that the issue had come up before the Civil War, but had stopped there, either not fully appreciating the breadth and duration of activism on this issue—how could they given the lack of any centralized measure or systematic source on the issue—or taking it as a background condition that could effectively be ignored in studying the debates over black suffrage post-1865.

The problem, from a causal analysis perspective, was that while this certainly did not invalidate dominant accounts of post-War black enfranchisement—that it was motivated by the strategic desire of the Republican Party to stay in power—it perhaps meant that we were overestimating the significance of this relative to the other accounts that stressed a more programmatic and social movement-based motivation (Cox and Cox 1967). If, as I was gradually discovering, a majority or near-majority of the Republican Party was willing to back black suffrage as early as 1860, then the motivations and strategic choices made by party leaders and issue entrepreneurs in 1867 would be considerably different. Instead of having to build support among a majority of their own party, they had to focus on the marginal Republicans, those often originating in former slaveholding states such as Missouri and West Virginia, whose votes would be needed to override President Johnson's veto. Appeals to party and stressing the threat of a revitalized Democratic Party were certainly important; but local activism, as it unfolded over the long term, meant that there was already a sympathetic and supportive base of Republicans to build on. Evaluating this dynamic process statistically and in a causal inference framework, however, would be next to impossible given the data that did not exist.

The more immediate and consuming problem from a research approach, then, was that because I initially lacked almost any data on the pre-Civil War efforts to remove racial qualifications for voting rights, the empirical research portion of my efforts had to be directed toward collecting and compiling this data, simultaneously building both a measure of the outcome as well as measures of the possibly relevant independent and confounding variables. I only gradually came to appreciate

that I needed to study the emergence of activism on this issue and legislative voting as a dynamic process, and not simply as an iteration of distinct and independent cases. But even if I had realized this at the outset I would have been largely without guidance in how I might accomplish this in a causal inference framework, given the inability to recover the type of data that would require.

Two tradeoffs inevitably present themselves. While I was busy compiling whatever information I could find—speeches, votes, state legislative district voting patterns and demographics, party affiliation, archival documents, referenda results, and secondary accounts—and even reducing much of this to quantitative form, I was ultimately going to produce a dataset in which most of the effort had gone into measuring the outcome (legislator votes on black suffrage), at the cost of right-side variables that might help me make a case that I had a fully specified model that had blocked out all "back-door" pathways to black suffrage (Morgan and Winship 2007; Pearl 1995). And while I could help us learn more about the antebellum era than we had previously known, the question and research design that I had begun with receded gradually toward the horizon.

In any case, what became clear in preparing the data and manuscript was that the dynamic nature of the process was going to be impossible to study in a statistical causal inference framework tailored to the problem of dynamic processes (such as Blackwell 2012); the data that was required to meet the assumptions of a causal inference design simply did not exist in a ready-to-use form and, as I came to believe, could not be made to exist in any form that would satisfy the requirements of being sufficiently granular, measured at enough points in the process, and pretreatment.

For this reason, the claims made in my study of black suffrage politics as it developed in the antebellum era (n.d.) have remained largely descriptive, and the gestures I make toward a more causal story come primarily from a qualitative tracing of the process by which black suffrage was put on the agenda and voted on across different states, supplemented where possible by descriptive statistical and quantitative data. This is a limitation of the study, in a period where well-identified causal inference is increasingly the gold standard; still, I believe that the empirical description that I offer there of how black suffrage developed as an issue over time is a valuable contribution to the study of American political development. And I suggest that it has at least the advantage of knowing its limits: recognizing the enfranchisement of African American men in the 19$^{th}$ century as a dynamic process makes it difficult to conceive of how we might study it as an event in a causal inference framework, and any single statistical estimate of the different potential causes at the moment of its

"final" occurrence will be biased unless it integrates this longer process. Studying the relationship of different variables at different times, and tracing their sequential unfolding, however, will leave us without a cleanly identified estimate of causal effect. In such a case, at least for the moment, descriptive statistical claims with causally oriented qualitative process tracing might be the best that can be made.

*Teele: rich data terrain meant expanding measures on the right hand side.*

For DLT, on the other hand, there was a ton of information on the women's suffrage campaign, particularly about the United States, that served as a jumping off point for statistical research. The suffragists themselves were keenly aware of the history-making they were involved in, and kept detailed records – published in six volumes as the *History of Woman Suffrage* – which they marched over to the Library of Congress shortly after the Nineteenth Amendment was secured in 1920. The often-repeated quote is that to win women the vote it took

> fifty-six campaigns of referenda to male voters; 480 campaigns to urge Legislatures to submit suffrage amendments to voters; 47 campaigns to induce State constitutional conventions to write woman suffrage into State constitutions; 277 campaigns to persuade State party conventions to include woman suffrage planks; 30 campaigns to urge presidential party conventions to adopt woman suffrage planks in party platforms, and 19 campaigns with 19 successive Congresses.

Each of these campaigns was recorded to a greater or lesser extent by the suffragists in the *History of Woman Suffrage*. When I first "discovered" these volumes for myself, my intuition was to try to skim them all and then code the whole thing up. Instead, I familiarized myself with the secondary literature on women's suffrage, and realized that much more data-informed work had been done on this topic than I had previously imagined.

Since we already knew the dates in which different states had allowed women "full" suffrage on equal terms as men, many scholars invested time in measuring causal variables. Lee Ann Banaszak (1996) plumbed the minutes of the National Woman Suffrage Association and came up with a measure (albeit imperfect) of membership in state level NAWSA branches over time. (We still lack local branch membership information, however.) Holly McCammon and her many coauthors had already examined the types of tactics that suffragists used over time, e.g. direct action tactics like protests, or writing newspaper editorials. Her team also considered the types of arguments suffragists made, whether appealing to the political expedience of enfranchising women or to justice

based arguments.[23] The findings that these scholars presented were convincing that the movement strategy mattered, but I hoped to do better on two fronts: first I wanted to know more about the larger history within states – why, for example, some states like Massachusetts repeatedly considered the issue in its legislature but was extremely resistant to adoption – and I wanted to know more about the political catalysts or hindrances.

My qualitative research into the suffrage politics in the United Kingdom (Teele 2014) convinced me that ruptures in political competition – whether you want to call it re-alignment or an increase in competition – led to entrepreneurial thinking on the parts of politicians. Although there were multiple legislatures in the whole of Great Britain that had some jurisdiction over suffrage (e.g. the Tynwald in the Isle of Man), my main concern was why the Westminster parliamentarians had resisted suffrage, even when the Liberals came to power in 1906, but ultimately enfranchised (most) women in 1918. I examined the debates and internal correspondence between legislators and suffragists in the period from 1910-1918 and constructed a timeline. The relevant counterfactual nodes were obvious: they were dates that private member bills for suffrage were proposed, debated, or voted on in each period. I studied the vote histories of the various parties (and the factions within the parties) and came to the conclusion that the changing political conditions that emerged in the wake of the First World War were instructive for suffrage only insofar as they increased competition, not because they changed anyone's minds (earlier bills with the same group of legislators had already reached a majority). The political holdouts continued to be a group of conservatives that were afraid of the leftism among women (so they only agreed to let older and wealthier women vote) and the leader of the Liberal party who also feared the direction of women's votes. The point is, that I came to the US case with a "prior" belief about the conditions under which suffrage would emerge and wanted to use the larger amount of variation across the US states as a plausibility probe or "hoop test" for examining my ideas.

The problem was, the fine-grained knowledge I had of the sequence of reform in the UK was not going to be possible to gain over 45-48 odd states. So instead I read as much of the historiography of the suffrage movement as I could, and I also delved into the literature on American Political Development (thanks in part to a reading list David Mayhew). Reading these works I began to form an intuition that competition was probably also instructive for the passage of

[23] They found that expedience-type arguments were more successful and may have accounted for early western extensions. Although during the course of my work I had asked McCammon to share her data, she was not able or willing to do so.

suffrage in the US, but that it operated in slightly different ways related not only to which party was in power, but how long they had held control, whether the power was projected across multiple levels of state government, and how big a majority they had. There was also the issue of political machines. Many suffragists felt that the machines were against women's enfranchisement because of the moral project that was a part of the suffrage movement. If women wanted to clean up dirty corrupt politics, the machines would be an obvious adversary. My first attempt at studying state-level enfranchisement quantitatively was then to try to measure competition in a more sophisticated way, and to replicate others' studies of the final reform. In 2011 I spent the year working on measuring political competition, and then I spent the summer thereafter, working with an RA, to collect information on political machines at the city level throughout the Gilded Age.

My first attempt at studying state-level enfranchisement quantitatively was thus to replicate others' studies of the final reform but to improve on the causal variables related to political competition. This was unsatisfying because the final date of enfranchisement was often not related to legislative passage within a state, but instead to a successful referendum. My intuition was that examining bill passage in the US, like I did in the UK, would allow for more statistical leverage than previous studies. But at some point in 2011 I discovered a gold mine: King, Cornwall, and Dahlin's (2005) research into the step-by-step legislative process that the suffragists had to overcome. This team of researchers had, amazingly, collected a database that listed (or approached) every single bill presented in every state legislature related to women's suffrage throughout the entire campaign. Their spreadsheet contained 1124 rows, 610 of which pertained to full suffrage rights, 562 of which were unique. Gaining access to their data felt less like standing on shoulders than soaring through the air. Like earlier scholars, their own research with this data had focused more on the social movement side (using some of McCammon's data in their analyses) and less on the political variables. Thus there was definitely room to contribute to the conversation.

The fact that data related to the historical sequencing of suffrage in the US states already existed was a huge boon to my project. To be sure, I had to do a lot of work to understand what was there (as the long appendix in Teele 2018b details), and there were certain things that I wish were different. For example, King et al. (2005) used the History of Women's Suffrage, supplemented by states' "Blue Books" to track the language of the bills and whether they voted on and whether, if so, they passed. When the information was available they recorded the name of the bill proposer but never the party, and bill passage was coded as a binary, not as a roll-call total. These are key pieces of information that I would have liked to know, both because I think that the partisanship of the

proposer vis-à-vis the partisanship of the legislature can provide insights into how power promotes or suppresses reform efforts, and because I would have liked to know whether the vote tallies followed a similar almost disjoint process as elsewhere – with small levels of support morphing almost seamlessly into supermajority levels (an S-shape adoption curve, if you will). These questions remain unresolved.

Ultimately I integrated something like studying the timeline of counterfactual nodes in two ways. First, by carefully reading through and cleaning all the data on bill proposals in state houses I became familiar with the protocols within states. I thought long and hard about what the correct unit of analysis was. Typically, scholars study legislative changes by using years as the unit of analysis. In a panel analysis of US state level reform, each state and each year will be its own row of the data. What I realized, though, is that often times the relevant legislative variables would not change across years (or we would not have measures of their changes) because legislatures did not turn over every year. This means that scholars that used a fixed-effects framework on state level panel data might estimate correlations between competition and suffrage that were downward biased precisely because the competition variables changed less frequently than other variables. In other words, fixed effects regressions look for how changes within variables within states are related to changes in the dependent variable. If certain variables do not change within years, this can attenuate coefficient sizes even though, in reality, whether the bill passes in the first year or the second of the legislative session may be less important than that it passed in that particular session. There is also the worry that off-the-shelf measures of competition are recorded at some point in the year that are not reflective of the actual composition of the legislature sitting when the bill was proposed. The lack of variation across years in legislative composition, and the potential to introduce post-treatment bias if the legislative composition was measured at the wrong point in time, caused a big headache. After taking a deep dive into various electoral histories to learn about when legislatures turned over in the Nineteenth Century, I decided to use the legislative session as the unit of analysis.

In addition, I decided to examine statistically the various stages of bill proposal and passage using a host of measures about the social movement and about political competition. I discovered that bills were not more likely to be proposed in years when competition was higher within states, but that the more competitive states considered more bills for suffrage overall. Looking at changes in political competition surrounding bills that were voted on, I discovered that states that became more competitive in the legislative session where a bill was voted on were more likely to pass the bill relative to states that voted on bills but which did not see higher levels of competition (some of this

is presented in Teele 2018a, ch 4). Finally, in what became the foundation of a stand-alone-article, I examined support for suffrage in state legislatures over time by analyzing all times that suffrage bills passed.

For the most part the models were not that sophisticated but instead the innovations came from conceptualizing competition in a more holistic fashion. I found robust evidence for the notion that competition was linked to greater support for suffrage, but that consolidated political power, whether in the form of longevity of the ruling party, the presence of political machines, and a larger majority surplus, was associated with resistance to suffrage. These relationships were, moreover, more profound in states where the suffrage movement was stronger, providing some evidence that competition impacted the efficacy of the suffrage movement.

*If we knew in graduate school what we know now:*

One big challenge to doing historical causal inference – which as we have argued requires thinking big *and* thinking small – is how to allocate time and resources. In the ideal world, the same person who does the burrowing will also recreate all the timelines, so that no juicy detail goes unnoticed. This is of course not feasible if the timelines are long, there are many nodes, or when there are many observational units (all of which we want for thinking big!). Inevitably, some research assistance will be a very useful thing, and we have a few suggestions for how to maximize the budget.

First, you have to do your thinking for yourself, so the most important thing is to have actually formed causal process observations for a good set of cases on your own. This is much easier to do if you work in a single language, but don't be sanguine that reading (much less locating) debates, blue books, and newspaper sources is easy even in the United States. Just start with a few cases (perhaps one that you already know something about, and two others that you know very little about) and dig in. In addition to helping you learn about the actual politics on the ground, you can begin to construct a coding scheme that someone else can follow later.

With actual substantive knowledge of particular timelines, and with a coding scheme in hand, you can apply for grants. In order to secure money, you need to have an interesting topic, a great hook, and to demonstrate that your project is feasible. Hence you need to actually have done some work in order to get money to do more. Having a sense of what data you hope to collect for the

universe of cases, and how difficult this will be, will allow you to estimate what it will cost to have helpers dig into other cases.

In addition, we want to argue that it is too high a hurdle to expect all new papers to create new data and new theory. One of the comparative disadvantages of political science when it comes to doing this type of work is that we don't slot graduate students around historical knowledge areas. So it means that we don't have good mentorship networks and that the infrastructure often does not get passed on. Sharing data with others, and acknowledging how big the playing field can be will allow for deeper understanding of major historical events. Since we are scientists, we must avoid the notion that there is a definitive book on anything, and make room for more friends in the sandbox. Finally, we believe that we need to value descriptive work, both because context-specific knowledge is key to causal inference, but also because measuring concepts creatively is hugely important to what we do.

## TOWARD A DEVELOPMENTAL PERSPECTIVE ON CAUSALITY

Scholars of political development in both comparative politics and American politics have increasingly begun to use quantitative tools to study the long-term persistence of institutions, to intervene in historiographical debates abut the causes of major institutional transformations, and to use historical examples as empirical cases in debates about general theories of politics. An enticing aspect of this body of research has been its engagement with and use of cutting edge techniques of design-based inference. Yet we argue that design-based techniques present problems for historically oriented social scientists when they seek to study major institutional transformations for the reason that particular realizations of relevant causal factors – such as the level of political competition or the ideological preferences of legislature – have qualities that act like both stocks and flows: they can operate on the outcome through both inherited and instantaneous pathways.

One solution to this problem, what Blackwell has called "dynamic causal inference", is promising albeit likely limiting both because of the opacity of the method and the lack of good data. Instead, we suggest that if we pay attention to historical sequencing, then we can use more familiar quantitative methods to examine the relationship between hypothesized causal variables and outcomes of interest over time. The claim that we need an understanding of historical sequencing for causal analysis of the politics of the past is unoriginal, as the large literature on historical institutionalism can attest (for a review, see Thelen 1999). Yet we hope that our modest addition –

that we must to reconstruct the timeline of relevant counterfactual moments – has other advantages as well, possibly even just in expanding our descriptive understanding of historical phenomena.

Perhaps most important, we should be careful not to assign causal inference a unique role in social scientific research. Indeed, we are all entirely familiar with different modes of demonstration and persuasion. The value of causal inference is enormous; but identifying a cause is only a subset of the more common practice of making a case, and not all cases rest on a dispositive identification of a causal effect. Modeling selection into treatment is itself an extremely difficult and complex task, worthy of entire areas of specialization. And while there might seem to be a historical literature on everything under the sun, it is very frequently the case that the literature that might inform the process of selection into treatment or help validate an instrument is sparse or non-existent. Insofar as this is the case, the plausibility of the "as-if" random argument often rests on marshaling evidence that the instrument was independent of whatever set of plausibly relevant variables the researcher is able to dig up. Lacking a literature, or deep knowledge of a literature, to help evaluate whether the demonstration of the assumption's plausibility, the reader is effectively asked to make a leap of faith – and as with any leaps, some are able and willing to jump further while others will refuse to leap at all. Thus we agree with others that if quantitative causal inference can only proceed if we believe in the soundness of qualitative inferences about the data generation process, quantitative studies of political development rests on the work of historians and historical institutionalists (Kocher and Montiero 2016: 956).

In closing, we want to make a call for what we are calling a developmental perspective on causality. We see this as an orienting practice common to many students of American political development and historical institutionalism, although it is not often framed as such nor articulated as a deliberate strategy for understanding causation. At its core are a set of simple premises, ones we expect most APD and historical institutionalist scholars would agree with but which are often backgrounded in our writing (even as they are often absent from work that treats history as data or that seeks to leverage temporality to make causal claims about the legacy of historic institutions).

These can be stated simply enough. (1) Sequence and timing matters, i.e., a static configuration of variables will not sufficiently explain outcomes if the sequence in which those variables appeared or occurred matters and that *when* an action occurs can be as important as *if* it occurred (Pierson 2004). (2) The relationships between variables and processes of interest will differ across distinct, but often hard to identify, temporal periods (Wawro and Katznelson 2014). This is a

corollary of the argument about timing, but concerns the different relationships between variables conditional upon the different parameters that accompany distinct institutional or political eras. In short, there are few transhistorical concepts or relationships, and a temporal bounding is essential to understanding the guidance and applicability of established theories or empirical relationships. (3) And finally, the configuration of variables that might be observable at the moment of an event's "occurrence" does not always provide an adequate explanation of their relative contributions over time.

A developmental perspective simply requires researchers to ask whether, to what extent, and to what consequence the events or processes that they are interested in are characterized by affirmative answers to these premises. Obviously, some will and some won't, and we might not know until we have actually set out to do the research; the point is not that everything is "development," but rather that researchers would do well to ask themselves at the outset about whether what they are studying is characterized by these dynamics and to plan the research strategy accordingly. Our concluding proposal, then, is to simply cultivate the habit of asking these questions of our research proposals, and to take seriously the difficulties – in terms of research design, data availability, and strategies for causal inference – for answering historical questions.

## REFERENCES


Abramson, Scott F., and David B. Carter. "The historical origins of territorial disputes." *American Political Science Review* 110, no. 4 (2016): 675-698.

Acharya, Avidit, Matthew Blackwell, and Maya Sen. *Deep Roots: How Slavery Still Shapes Southern Politics*. Vol. 6. Princeton University Press, 2018.

Banaszak, L. A., 1996. *Why Movements Succeed or Fail: Opportunity, Culture, and the Struggle for Women's Suffrage*. Princeton University Press.

Bateman, David A. 2018. *Disenfranchising Democracy: Constructing the Electorate in the United States, the United Kingdom, and France*. Cambridge University Press.

Bateman, David A. n.d. "The Development of Party Positions on Black Suffrage." Unpublished Working Paper.

Bateman, David A., Ira Katznelson, and John Lapinski. 2018. *Southern Nation: Congress and White Supremacy After Reconstruction*. Princeton: Princeton University Press.

Bensel, Richard Franklin. 2009. *The American Ballot Box in the Mid-Nineteenth Century*. Cambridge: Cambridge University Press.

Blackwell, Matthew. "A framework for dynamic causal inference in political science." *American Journal of Political Science* 57, no. 2 (2013): 504-520.

Blight, David W. 2018. *Frederick Douglass: Prophet of Freedom*. New York: Simon & Schuster

Carus, A.W., and Sheilagh Ogilvie. 2009. "Turning Qualitative into Quantitative Evidence: A Well-Used Method Made Explicit." *The Economic History Review* 62(4): 893-925

Cirone, Alexandra, and Brenda Van Coppenolle. "Cabinets, Committees, and Careers: The Causal Effect of Committee Service." *The Journal of Politics* 80, no. 3 (2018).

Cox, LaWanda, and John H. Cox. 1967. "Negro Suffrage and Republican Politics: The Problem of Motivation in Reconstruction Historiography." *The Journal of Southern History*. 33(3): 303-330

Eggers, Andrew C., and Arthur Spirling. "Incumbency effects and the strength of party preferences: Evidence from multiparty elections in the United Kingdom." *The Journal of Politics* 79, no. 3 (2017): 903-920.

{Falleti Decentrailization Essay}

{Falleti and Mahoney Advances article}

Field, Phyllis F. 1982. *The Politics of Race in New York: The Struggle for Black Suffrage in the Civil War Era.* Ithaca, NY: Cornell University Press

Gaikwad, Nikhar. "East India Companies and Long Term Economic Change in India." *American Political Science Association.* Unpublished manuscript.

Gerring, John. 2012. *Case Study Research: Principles and Practices*. Cambridge: Cambridge University Press.

{Gryzmala Busse Time Will Tell. }

Haggard, S., Kaufman, R.R., 2012. "Inequality and Regime Change: Democratic Transitions and the Stability of Democratic Rule." *American Political Science Review* 106: 495–516.

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

Jusko, Karen Long. *Who speaks for the poor?: electoral geography, party entry, and representation*. Cambridge University Press, 2017.

King, B. G., Cornwall, M., Dahlin, E. C., 2005. "Winning Woman Suffrage One Step at a Time: Social Movements and the Logic of the Legislative Process." *Social Forces* 83: 1211–1234.

Kocher, Matthew A., and Nuno P. Monteiro. "Lines of Demarcation: Causation, Design-Based Inference, and Historical Research." *Perspectives on Politics* 14, no. 4 (2016): 952-975.

Lustick, Ian S. 1996. "History, Historiography, and Political Science: Multiple Historical Records and the Problem of Selection Bias." *American Political Science Review*. 90(3): 605-18

McConnaughy, C., 2013. *The Woman Suffrage Movement in America: A Reassessment*. Cambridge University Press.

Mahoney, James. "Strategies of Causal Inference in Small-N Analysis." *Sociological Methods and Research* 28:4 (2000), pp. 387-424.

Morgan, Stephen L., and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge: Cambridge University Press

Neyman, Jerzy. 1923 [1990]. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9." Dorota M. Dabrowska and Terence P. Speed, trans. *Statistical Science* 5 (4): 465–472.

Pearl, Judea. 1995. "Causal Diagrams for Empirical Research." *Biometrika*. 82(1): 669-88

Pierson, Paul. 2004. *Politics in Time: History, Institutions, and Social Analysis*. Princeton: Princeton University Press.

Rohlfing, Ingo. 2012. *Case Studies and Causal Inference: An Integrative Framework*. New York: Palgrave Macmillan

Rothbauer, Paulette M. 2008. "Triangulation." *SAGE Encylopedia of Qualitative Research Methods*. Thousand Oaks, CA: SAGE, 893-95

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies." *Journal of Educational Psychology*. 66(5): 688-701

Simon, Joshua. 2014. "The Americas' More Perfect Unions: New Institutional Insights from Comparative Political Theory", Perspectives on Politics, volume 12, number 4 (December 2014), 808-828.

Teele, Dawn (ed.). *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experiments in the Social Sciences*, Yale University Press. 2014.

Teele, Dawn Langan. 2018a. *Forging the Franchise: The Political Origins of the Women's Vote*. Princeton University Press.

Teele, Dawn Langan. 2018b. "How the West Was Won: Competition, Mobilization, and Women's Enfranchisement in the United States." *The Journal of Politics* 80(2): 442-461.

Thelen, Kathleen. "Historical institutionalism in comparative politics." *Annual review of political science* 2, no. 1 (1999): 369-404.

Vitalis, Robert. 2006. The past is another country. *A Handbook for Social Science Field Research: Essays & Bibliographic Sources on Research Design and Methods (Ellen Perecman and Sara R. Curran, eds.)*, pp.5-17.

Voigtländer, Nico, and Hans-Joachim Voth. "Persecution perpetuated: the medieval origins of anti-Semitic violence in Nazi Germany." *The Quarterly Journal of Economics* 127, no. 3 (2012): 1339-1392.

Wang, Xi. 1997. *The Trial of Democracy: Black Suffrage and Northern Republicans, 1860-1910*. Athens, GA: University of Georgia Press.

Wawro, Gregory J., and Ira Katznelson. 2014. "Designing Historical Social Inquiry: How Parameter Heterogeneity Can Bridge the Methodological Divide between Quantitative and Qualitative Approaches." *American Journal of Political Science*. 58(2): 526-46

Weaver, Michael. 2019. "'Judge Lynch' in the Court of Public Opinion: Publicity and De-legitimation of Lynching." (Forthcoming, *American Political Science Review*).