A Developmental Approach to Historical Causal Inference

David Alexander Bateman, Cornell University

Dawn Langan Teele, University of Pennsylvania[1]

ABSTRACT: Empirical historical research typically falls into one of three types: the study of major historical events; the use of "history as data" to test general theories; and the study of contemporary legacies of historical processes. We argue that because of data sparsity and dynamically unfolding processes, the study of major historical events is less well suited to design-based inference than other types of historical research. Drawing examples from our own work, we suggest that the construction of a "timeline of relevant counterfactual nodes" can structure thinking about complex historical processes. The researcher can focus on these relevant counterfactual moments as potential episodes of reform using either statistical or qualitative techniques as appropriate, moving forward through the timeline and updating their beliefs about a hypothesized cause's importance across the process.

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 what causation is, and 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; Splawa-Neyman 1990) – there has been a renewed appreciation for, and use of, experimental and so-called quasi-experimental methods across the discipline (Dunning 2012; Morgan and Winship 2007; Pearl 1995). 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 single standard for conducting research into the present or the past (Sekhon and Titiunik 2012). Its proliferation, while rapid, has been uneven, both because the design-based inference approach is more closely aligned with narratives of "single" causation, and because it is easier to carry out in a knowledge terrain that is already rich with historiographical debates and prior scholars' efforts to collect data.

Historical social scientific questions that lend themselves to empirically grounded causal analyses come in different types, three of which are particularly common in political science and cognate

---

[2] On the historical turn see Capoccia and Ziblatt 2010; on the relationship between American political development and comparative politics see Morgan 2016. On designed-based approaches see Imbens 2010, 403; Keele 2015, Dunning 2012, and Diamond and Robinson 2010.

fields. 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 data" 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 the 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.

Our contention is that the strictures of most 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 both the historical legacies and the history-as-data genres 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 these modes can even supplement historical data with newly constructed data that they can tailor to better meet the assumptions of design-based inference.

---

[3] We thank Scott Abramson for sharing these insights at the historical political economy working group meeting, APSA 2017. Examples of historical legacies research include AJR 2001; Voigtlander & Voth 2012; Nunn 2008; Acharya, Blackwell, and Sen 2018; Abramson and Carter 2016.

[4] This mode of historical research does not necessarily imply stable causal relations across different temporal periods, nor does it ignore changing temporal contexts. Indeed, much research in this vein is explicitly tailored to examining whether theoretical foundations that have an established basis in one period apply in another; if they do not, the explanation for this can serve as a valuable refinement of theory. E.g. Eggers and Spirling 2017; Cirone and Van Coppenolle 2018; Cox et al. 2016; Weaver 2019.

The third type of historical research, on particular historical events, is arguably one of the most common type of question asked by scholars working in American political development, and we suggest that it is also the most difficult to study in a design-based framework. Questions about why a particular event occurred 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 severe 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, then modeling the causal process requires close attention to the possibility for co-determination in the values of important co-variates.

Our paper details some of the common challenges that confront researchers looking to make causal arguments in the third of these types of questions. 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 identifying a causal effect for a single variable than about evaluating the relative contribution of different theorized causes whose importance and effects might develop dynamically over an extended period of time. Even when the framework is explicitly comparative, the resulting research often veers toward being more case-centered than theory-centered, i.e., it is focused on providing a fuller accounting of a few

specific events or processes rather than empirically establishing a generalized theory of how different variables or processes relate to each other across cases (Rohlfing 2012).[5] And because we are interested in explaining a particular case or set of cases, we cannot always assume that observing and recording data from similar phenomena in the future would provide much additional leverage on explaining the past, even if we were to make the often ahistorical assumption that a causal relationship established for one period applies with equal force in another (Haydu 1998). What is more, where data repositories exist or could be recovered, they are almost certain to have been generated and collected with different ends in mind; even with a deep dive into the archives and original data collection, the empirical foundations of the analyses will generally be sparser and more conceptually tangential than we would desire.

These characteristics of much historical research – a concern with particular cases whose causes unfolded dynamically and for which we have often very limited data – present inferential difficulties that cannot be mechanically addressed within a potential outcomes framework. We suggest instead that a more pressing need for establishing causal inference is a set of procedures for guiding historical researchers that will facilitate recognizing dynamic causal processes where they occur, and which will help in the logical evaluation of how well different causal hypotheses hold up across the totality of a developing historical process.

We begin with short vignettes describing the research questions, processes, and outcomes from our recent work on franchise expansion to black men and women. These vignettes help to ground three claims we advance in the remainder of the paper. The first is simply that prominent ways of analyzing historical events – fixed effects regression frameworks, event history analysis, but also

---

[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.

many of the logical tests researchers use to evaluate qualitative process tracing studies – are likely to produce biased estimates of hypothesized causal factors if they fail to model their developmental and dynamic character across a particular period. This will be true whether our objective is estimating a causal effect across multiple cases or establishing the validity or weight of a causal hypothesis in a single case. As a result, studying particular historical events should, as a matter of general practice, integrate some features of process-tracing, even if just to establish the appropriateness of a design-based inferential model. Good statistical work on historical events, like good qualitative work, will require scholars to be close to their cases, regardless of the epistemological grounding of their approach (Goertz and Mahoney 2012; Kocher and Monteiro 2016, Vitalis 2006, Lustick 1996).

The second claim follows from the first, and is the foundation for the positive contribution we hope to make with this article. Because of the difficulties of studying dynamic processes, especially given the data sparsity and unevenness ubiquitous to much historical research, there will not be a mechanical solution that can be applied generally (e.g. like that described by Blackwell 2012). Instead, those of us interested in making causal inferences about historical events should focus on developing a set of research procedures that can be used to guide research design and facilitate causal inference across a historically unfolding process. In particular, we propose that scholars center their data collection and analysis around a timeline of "relevant" counterfactual moments—temporally defined instances in which an outcome actually was possible but did not occur (Simon 2014). The researcher can then study these moments using either statistical or qualitative techniques as appropriate, moving forward through the timeline and updating their beliefs about a hypothesized cause's importance across the process.

Our third claim is simply that as a discipline our valuation of historical research should be relative to the prior state of knowledge. Statistical estimation of a causal effect is not the only means by which causal inference can be undertaken – given sufficiently specified theory, description itself is a powerful tool for establishing causality (Falleti 2016) – and in the study of historical events it will often be impossible within a single unified framework. In a terrain with deep prior knowledge and rich data sources, scholars may be able to make advances using quantitative concepts of identification; in sparser fields, by contrast, the descriptive documentation of a process, including the recovery and analysis of possible counterfactual moments, is itself an important contribution to knowledge and causal inference (Kreuzer 2019). We suggest in the conclusion that these three claims amount to taking a developmental perspective on the conduct and evaluation of historical research, one that is sensitive to temporality and sequencing, that facilitates causal inference across an unfolding process, and which does not prioritize estimation of a single causal effect over other more cumulative contributions to knowledge.

QUESTIONS, PROCESSES, AND OUTCOMES

In the process of conducting our separate research into black suffrage and the antecedents of women's enfranchisement, we began to understand that our first intuitions about how to study the passage of suffrage reforms failed to appreciate or adequately model the sequencing of the historical process that led to reform. Moreover, while we both tackled problems of compiling and measuring scarce data for what we eventually came to see as dynamic processes, we confronted very different existing research infrastructures in terms of the prior knowledge and data sources we had readily available. This had big implications for the allocation of our time in the research process, and for the resulting products of our work.

Bateman: sparse data and the tradeoffs of research

Bateman's plan for a study on African American voting rights began with a rough timeline of suffrage reform across American history. The basic trajectory I was interested in is shown in Figure 1 (Bateman forthcoming), which tracks the growing proportion of states that disenfranchised African Americans. The initial research strategy was to compare the politics of black voting rights across three periods, beginning with the first wave of disenfranchisement that ended in the mid-1830s, then turning to the mass enfranchisements of 1866 to 1870, and concluding with the mass disenfranchisements of the late 19th and early 20th century. I anticipated the bulk of my research time to be spent studying the second and third periods.

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, but that the period I was planning on skipping over entirely – from approximately 1835 to 1865, where Figure 1 shows nothing much happening – had in fact seen a considerable amount of political activity, including speeches, writing, legislative voting, public referenda, and petitioning on the issue of black suffrage. An issue that exploded onto the political agenda in 1867, and remained at a fever pitch for years afterwards, had already been a feature of political debate and the subject of intense legislative fights at the state level for at least two and a half decades.

Certainly, black suffrage before the Civil War was a different thing than it would become afterwards: enfranchising the relatively small free African American population in any given state would not have been the same as expanding the right to vote to millions of newly freed men in the

South. But many of the arguments being deployed in 1867 for and against black suffrage had already been made; numerous states had already experienced intense political organizing on this issue; 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 (Field 1982).
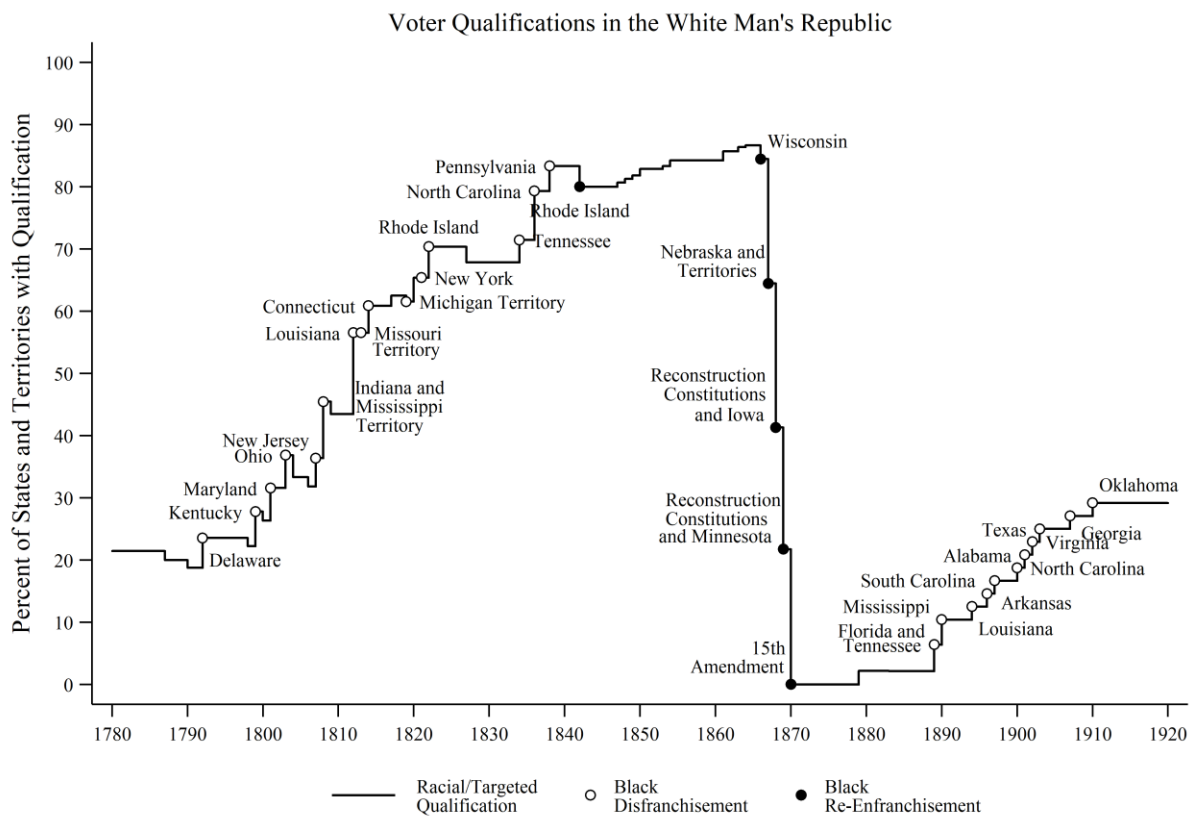

Figure 1: Timeline of Reform on Racial Voter Qualifications. Outcome rather than Process-Focused.

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 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 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 (Valelly 2004)—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 (eg., Cox and Cox 1967), rooted in both the extraordinary organizing of newly free persons described by Valelly but also the older antislavery societies and conventions of free persons of color. If large factions of the Republican Party had been willing to back black suffrage as early as 1860, then the motivations and strategic choices made by party leaders and issue entrepreneurs in 1867 took on a different cast. Appeals to party and arguments 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. Instead of having to build support among a majority of their own party, party leaders had to focus on the pivotal Republicans, often from border states or states where abolitionists had not been effectively organized. In short, I had come to believe that the earlier activism had dynamically altered the state of public opinion and the relative importance of electoral calculations to the final outcome: instead of studying these as a succession of independent cases, they had to be considered in a temporal sequence.

There is, however, little guidance on how to study a dynamic causal process in a potential outcomes framework, and that which exists (Blackwell 2012) requires data at a level of temporal and spatial granularity that simply did not exist. Moreover, while I had already decided upon a mixed-methods approach, pairing qualitative process tracing with statistical evaluations of legislator voting, the data infrastructure for voting in state legislatures and constitutional conventions was next to non-existent for this period. Lacking an existing data repository for identifying when and where the issue was debated and voted on, the time and resources allocated to this task took priority over recovering conditioning variables necessary to construct a fully specified model blocking out all "back-door" pathways to black suffrage (Morgan and Winship 2007; Pearl 1995).

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. Recognizing the enfranchisement of African American men in the 19th 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 mountain 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.[6] 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.

---

[6] 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.

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 other moments of potential reform 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 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. (See table 1.) 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.

Table 1: All unique attempts at full women's suffrage bills in the US States, 1850-1920.

State	House Bills			Senate Bills			Ballot Measures				
	Introd.	Voted	Passed	Introd.	Voted	Passed	Initiat	Legisl.	Const	Type	Passed
AL	1	1	0	1	1	0					0
AR	5	3	2	4	3	3			1	C	0
AZ	11	10	7	12	6	1	1			I	1
CA	13	5	2	20	9	5		2		LL	1
CO	6	4	2	4	3	3		2		LL	1
CT	8	4	2	2	1	1					0
DE	4	3	0	5	3	0					0
FL	7	3	0	4	2	2					0
GA	8	0	0	3	0	0					0
IA	27	16	12	27	17	10		1		L	0
ID	5	4	1	2	1	1		1		L	1
IL	3	1	0	3	1	0					0
IN	12	5	4	10	6	3					0
KS	13	5	4	9	6	3		3		LL L	1
KY	1	1	0	2	1	0					0
LA	4	4	2	4	2	1		1		L	0

State										
MA	23	8	2	9	4	4		1		L	0
MD	5	0	0	3	1	1					0
ME	8	6	2	5	5	2		1		L	0
MI	14	9	4	7	6	3		4		LL LL	1
MN	8	3	3	9	3	0					0
MO	12	1	0	3	0	0	1			I	0
MS	3	3	2	5	4	2					0
MT	10	5	2	3	1	1		1		L	1
NC	2	0	0	2	0	0					0
ND	14	10	6	12	12	9		1		L	0
NE	7	4	2	7	3	1	1	1	1	IL C	0
NH	1	0	0	4	4	3			1	C	0
NJ	6	4	3	2	0	0		1		L	0
NM	1	0	0	12	9	6					0
NV	10	10	5	14	7	6		1		L	1
NY	28	12	5	6	2	1		2		LL	1
OH	7	3	1	8	1	1	1		1	IC	0
OK	10	4	4	8	7	4	1	1		IL	1
OR	10	8	6	9	3	3	4	2		IIII LL	1
PA	7	5	3	5	3	3		1		L	0
RI	6	5	3	4	3	0		1		L	0
SC	6	2	0	15	13	10					0
SD	14	12	8	2	1	1		6		LL LL LL	1
TN	1	1	1	3	1	1					0
TX	10	2	1	1	1	1		1		L	0
UT	1	1	1	1	0	0			1	C	1
VA	3	3	0	3	3	1					0
VT	1	1	1	7	6	5					0
WA	8	7	4	15	12	8		3		LL L	1
WI	26	16	6	7	4	1		1		L	0
WV	7	5	2					1		L	0
WY									1	C	1
Total	397	219	115	303	181	111	9	40	6		15

Note: Attempts to pass women's full-suffrage bills at the state level 1850-1920. The table lists the total number of bills introduced, voted on, and passed in the state house and state senate, as well as the total number of ballot measures, be they ballot initiatives, legislative referendum, or referendum on constitutional conventions that contained language for women's suffrage. Source: Teele's calculations from Cornwall's Suffrage Database.

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 relevant counterfactual moments 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.

THE FUNDAMENTAL PROBLEM OF HISTORICAL CAUSAL INFERENCE

Our studies of suffrage reforms highlight several difficulties for inferring causality in the study of specific historical events. The first, most straightforward, and ultimately most intractable is simply that the nature of the research questions makes it effectively impossible to generate new data that might meet the specifications of the potential outcomes framework. This difficulty is compounded by what emerged as a central characteristic of the events in question, which occurred not as a single occurrence but as the culmination of a developing and sequential process. Finally,

the existing research infrastructures that we could rely on as we assembled our empirical material and theorized our cases was of widely different quality, necessitating inevitable tradeoffs in the data that could be recovered or compiled and in the knowledge about cases that we could use to guide our resource allocation.

The fundamental problem of causal inference as defined in the potential outcomes framework 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 ever be observed. 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, and 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. Efforts to circumvent this difficulty in historical work have included the turn toward what is known as "design based inference," 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). Design-based approaches can include instrumental variable models, natural experiments (Dunning 2012, Sekhon and Titiunik 2012), the use of directed acyclic graphs to facilitate conditioning for so-called collider variables (Morgan and Winship 2007; Pearl 1995), regression discontinuity designs, difference-in-difference models, as well as efforts to empirically establish hypothesized causal relations using contemporaneously generated data, for example by using an experiment to establish individual-level mechanisms.

Design-based approaches are an extremely powerful tool, and have already done much to advance historical research. But as we suggested in the introduction, their utility is not distributed equally across different types of historical questions. Studies that attempt to explain contemporary cross-sectional phenomenon as the legacy of long-term processes, as well as those which use history as a source of data point to examine general theories, are perhaps especially compatible with the data assumptions and requirements of design-based approaches. By contrast, questions about the different causes of specific historic events—involving outcomes that are case-specific and have already occurred (Mahoney, Kimball, and Koivu 2009; Roberts 1996)—may be the least well suited to such an approach. 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 on the universe of cases to which a hypothesized theory might apply.[7]

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. While the researcher can and should draw on generalized theories, their applicability or weight in the particular instance of interest needs to be demonstrated rather than assumed; this very particularity is in turn likely to limit the types of data that the researcher will be able to bring to bear in making causal inferences. 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 is deliberately structured to better meet the standards required of design-based inference. 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. And where such an infrastructure exists, as it often does in well-plumbed research areas, then the likelihood of recovering new data that has not already contributed to, and could therefore serve as the basis for testing existing theories, is likely to be slim.[9] The researcher will then be required to engage in explicit modeling of an iterative research process, where the conditional probability of competing established theories being correct is evaluated through, for example, a Bayesian "dialogue with the data" (Fairfield and Charman 2019).

As we saw with the vignettes above, these intrinsic data limitations intersect with a more fundamental issue, which is how we conceptualize the relevant outcome in this type of analysis. As Paul Pierson (2004) 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. Following Pierson and Blackwell, we define a dynamic causal process as one in which the relevant actions or variables that influence an outcome of interest do not occur just once but as part of an

---

[7] 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.

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

[9] As Charles Ragin notes with regard to the "well-reasoned" argument that new or refined theories can only be tested on newly collected data, this effectively "puts an end to most case-oriented research" or where "the number of relevant cases is limited by the historical record to a mere handful" (Ragin 2004, 126).

unfolding sequence, in which the order 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 subsequent moments. Such processes pose considerable difficulties for causal inference, regardless of whether the mode of inquiry is statistical, qualitative process tracing, or a mix of different methodologies (Blackwell 2012). Since sequencing is often key for understanding historical events, analysis of event snapshots will often be hard pressed to meet the basic assumptions necessary for causal identification. 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.

Consider the two ways that historical events tend to be modeled in a statistical framework, both of which treat the occurrences in question as "single-shot" events. The first is to collect repeated cross-sections for specific observational units (legislatures, countries, organizations) and to measure a host of variables at proximate 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. Fixed effects regressions cannot estimate coefficients for time invariant variables (because the matrix cannot be inverted). Thus the inferential leap is made on the basis of relative changes in independent variables over time (a flow), rather than the stock of those variables. 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. Yet, 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. A cross-sectional or event history analyses of an occurrence at the moment of outcome, i.e., passage of reform, will show posttreatment bias in the estimated coefficients insofar as the values taken by key variables will be a function to their earlier exposure to other variables of interest.

A similar problem confronts qualitative historical researchers. Consider some of the different tests used in process-tracing, the "hoop test," "smoking-gun" test, or the promising exercise in Bayesian probability as "extended logic" outlined in Fairfield and Charman (2019). A hoop test simply proposes that "a given piece of evidence from within a case should be present for a hypothesis

to be true," and while it is seen as a necessary condition for establishing the validity of a given hypothesis it is also an insufficient one (Mahoney 2015, 207; Bennett 2008). A smoking-gun test, by contrast, asks whether there is evidence of a condition for which the cause or outcome in question is necessary; passing such a test is sufficient but not necessary to establishing the validity of a given hypothesis (Mahoney 2012, 574–78).[10] In each case, the validity of a hypothesis is established at a particular moment using the totality of evidence that is available. But as with "single-shot" statistical analyses, the arbitrary temporal bounding of the event will fundamentally alter the types of data that is available, as well as the estimated or assessed importance of different factors at the particular moment of its occurrence. If social movement organizing is relatively muted at the end moment of reform, for example, it might be interpreted as failure to pass a hoop test, despite having earlier played a role in creating the more favorable public opinion that was essential to passage. Smoking-gun evidence for a hypothesis might equally be lacking at the moment of reform, even though it might have mattered both directly and indirectly through its effects on other variables earlier in the process.

In one sense, the difficulties of modeling dynamic causal processes can be generalized to 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. The very generalizability of this problem means that there is nothing distinctly historical about it, and different solutions have been proposed for its resolution. These approaches generally on a large amount of fine-grained and temporally comprehensive data, which allows the researcher to estimate an integrated model that not only takes time into account but allows for development in the variables of interest (see also Wawro and Katznelson 2014). 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. Dynamic causal processes

---

[10] In the first test, the researcher asks whether all conditions defined by a particular hypothesis as necessary for an outcome's occurrence are observable in the historical record, or whether the "auxiliary traces" of a theoretically relevant latent concept are present. In the second test, the researcher asks whether a given hypothesis being true is a necessary condition for the presence of a particular phenomenon in the historical record.

are not something that can be reliably modeled in a statistical framework in much historical work, given the frequent sparsity of the data and, just as important, its temporal unevenness across the timeline of the process.

THE TIMELINE OF RELEVANT COUNTERFACTUALS

Given the intrinsic difficulties of making causal inferences in cases of dynamic causal processes we believe researchers would benefit from the formulation of a set of procedures that can facilitate the recognition of such a dynamic and the application, upon an evaluation of the likely data sources and limitations, of appropriate methods. The inevitable variation in the quality and coarseness of data, whether amenable to quantification or not, means that there will not be a mechanical solution that can be applied generally. Rather, what we suggest are a set of procedures intended to facilitate the empirical substantiation of causal arguments in cases of possible dynamic causal processes.

Our solution to the problem of studying dynamic processes that lead to major historical events requires thinking big and thinking small. Large-n researchers want to increase the number of cases that they study to be able to describe patterns and ultimately infer causal relationships over a large domain, be it spatial or temporal or both. Small-n researchers want to burrow into the details of the cases to have a rich understanding of the facts on the ground, that is, to be able to make more specific claims about causal relationships. Although some have suggested that the ontological orientations of researchers working in each of these modes may be different (Goertz and Mahoney 2012), we argue that it is possible to design a strategy for one's research agenda that does both, and that this type of design is crucial for studying major historical events quantitatively.

Specifically, we propose that the first step for researchers interested in understanding major historical events is to 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. A basic bias of studying historical events is that our interest is intrinsically backward-looking—we know the outcome, and are usually looking for antecedent causes—while it has long been recognized as a matter of historical practice that inquiry should "read history forward to reduce hindsight bias and conceptual reification" (Kreuzer 2019, 127; Skinner 1969). What we propose is that the researcher begin with a backward-looking construction of the sequence that plausibly led to an ultimate outcome, and use

this to structure a forward-looking narrative or sequenced analysis culminating in the occurrence or event of interest.

Starting with a timeline of potential counterfactuals entails re-conceptualizing the outcome not as a predetermined occurrence but as the product of a temporally developing set of factors, drawing out moments when the outcome could plausibly have occurred but did not. Constructing such a timeline will require us to 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 mean 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 change was plausible. This will in turn establish "nodes" in the history of an event in which the change could have occurred, providing information about relevant counterfactuals moments (Simon 2014). These counterfactual nodes are conceptualized in two distinct ways: as alternative outcomes in which the hypothesized causal and conditioning factors might have taken on distinct values—the standard basis for comparative analyses—but also as sequentially and developmentally linked moments, in which later nodes are likely to have developed out of earlier ones.

With such a timeline in hand, the researcher can then walk forward through the nodes, evaluating the strength of different hypotheses at sequential moments using the available or recoverable data and most-appropriate inferential approaches. This can involve statistical analyses, hoop tests, smoking-gun tests, or qualitatively "inhabiting the world of each hypothesis" and assigning probabilities about the truth value of empirical propositions conditional on these hypotheses and our background knowledge (Fairfield and Charman 2017, 369). At each moment, we rely on the available empirical evidence and logical evaluation to assess not 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. Given the inevitable unevenness in the data, the researcher's evaluation of the hypothesized relationships at any particular node is likely to be descriptive; but if rigorously and explicitly connected to theory, the researcher will be in a better position to make an inferential argument that rests not on any particular moment but across the totality of the causal sequence.

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 facilitate 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. Where the infrastructure is sparse, the researcher's main contribution might be to help construct it, a task that "requires sleuthing, language proficiency, familiarity with the organization of archives, knowledge about legal restrictions guiding their access, intuitions of what might have been deliberately omitted or destroyed, and above all, persistence" (Kreuzer 2019, 125). Still, the finding of new facts and their array into a coherent temporal sequence will build on the existing infrastructure. 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.

How we configure temporal and geographic variation across our timeline will shape the types of comparisons we make (Kreuzer 2019, 128), whether between temporally sequenced events or between spatially separate events treated as occurring simultaneously, as well as our allocation of resources. 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 generally provide variation along key dimensions of the dependent, treatment, and conditioning variables. 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.[11] 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. 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

---

[11] 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.

place, hence for thinking about how things like duration, tempo, and acceleration matter for an event's occurrence, 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 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 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.

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 because its requirement of temporally fine-grained 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.

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

Perhaps most important, we hope to underscore that casual inference, as understood in the potential-outcomes framework, should not be given pride of place in growing literature on historical

political economy. 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.[12] 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. (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 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

---

[12] By developmental we do not intend to imply anything about progress or movement toward a teleological end point. Rather, we use the term in the way it most commonly appears in the literature on American political development, to indicate historically-situated change from one temporal period or regime to another.

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 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.

Bennett, Andrew. 2008. "Process Tracing: A Bayesian Perspective." In The Oxford Handbook of Political Methodology, eds. Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier. Oxford: Oxford University Press, 217–70. https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199286546.001.0001/oxfordhb-9780199286546-e-30 (May 1, 2019).

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.

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, Gary W., Jon H. Fiva, and Daniel M. Smith. "The contraction effect: How proportional representation affects mobilization and turnout." The Journal of Politics 78, no. 4 (2016): 1249-1263.

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.

Diamond, Jared, and James A. Robinson, eds. Natural experiments of history. Harvard University Press, 2010.

Dunning, Thad. 2012. Natural Experiments in the Social Sciences: A Design-Based Approach. Cambridge: Cambridge University Press.

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.

Fairfield, Tasha, and Andrew Charman. 2019. "A Dialogue with the Data: The Bayesian Foundations of Iterative Research in Qualitative Social Science." Perspectives on Politics 17(1): 154–67.

Fairfield, Tasha, and Andrew E. Charman. 2017. "Explicit Bayesian Analysis for Process Tracing: Guidelines, Opportunities, and Caveats." Political Analysis 25(3): 363–80.

Falleti, Tulia G. 2013. "Decentralization in time: A process-tracing approach to federal dynamics of change." Federal Dynamics: Continuity, Change, and the Varieties of Federalism: 140-166.

Falleti, Tulia G., and James Mahoney. 2015. "The comparative sequential method." Advances in comparative-historical analysis. 211-239.

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.

Goertz, Gary, and James Mahoney. 2012. A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences. Princeton, N.J.: Princeton University Press.

Grzymala-Busse, Anna. "Time will tell? Temporality and the analysis of causal mechanisms and processes." Comparative Political Studies 44, no. 9 (2011): 1267-1297.

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.

Haydu, Jeffrey. 1998. "Making Use of the Past: Time Periods as Cases to Compare and as Sequences of Problem Solving." American Journal of Sociology 104(2): 339–71.

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.

Kreuzer, Marcus. 2019. "The Structure of Description: Evaluating Descriptive Inferences and Conceptualizations." Perspectives on Politics 17(1): 122–39.

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

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

——. 2012. "The Logic of Process Tracing Tests in the Social Sciences." Sociological Methods & Research 41(4): 570–97.

——. 2015. "Process Tracing and Historical Explanation." Security Studies 24(2): 200–218.

Mahoney, James, Erin Kimball, and Kendra L. Koivu. 2009. "The Logic of Historical Explanation in the Social Sciences." Comparative Political Studies 42(1): 114–46.

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.

Ragin, Charles C. 2004. "Turning the Tables: How Case-Oriented Research Challenges Variable-Oriented Research." In Rethinking Social Inquiry, eds. Henry E. Brady and David Collier. Lanham, MD: Rowman & Littlefield, 123–38.

Roberts, Clayton. 1996. The Logic of Historical Explanation. University Park, Pa.: Pennsylvania State University Press.

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

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.

Sekhon, Jasjeet S., and Rocio Titiunik. "When natural experiments are neither natural nor experiments." American Political Science Review 106, no. 1 (2012): 35-57.

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.

Skinner, Quentin. 1969. "Meaning and Understanding in the History of Ideas." History and Theory 8(1): 3–53.

Splawa-Neyman, Jerzy. 1990. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9." Statistical Science 5(4): 465–72.

Teele, Dawn. 2014. "Ordinary Democratization: The Electoral Strategy that Won British Women the Vote", Politics & Society, 42(4): 537-561.

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." American Political Science Review 13(2): 293-310.