Analyzing Natural Experiments:
A Comparative Historical Approach

Benjamin Smith*
University of Florida

With the proliferation of experimental methods in the social sciences and studies based on them have come a set of new challenges for both qualitatively and quantitatively oriented scholars. In this essay I develop an approach to natural experiment research based on comparative historical tools. By pairing such natural experiments of history with strategies from the comparative historical analytic (CHA) toolbox, it is feasible to demonstrate the "as-if-random" elements of treatment assignments drawing on the analytic parallels between critical junctures and treatment moments. It is further possible to adjudicate between multiple bundled treatments, and to generate and explore hypotheses about configurative treatment effects over time. I use the cases of two kinds of natural experiments of history—island colonization and ethnic region partition—to illustrate the specific benefits of wedding natural experiments with comparative historical analysis. Finally, I present a set of research strategies for incorporating natural historical experiments into multi-method research designs.

An early version of this paper was presented at the annual meetings of the American Political Science Association, August 28-31, 2014, Washington, DC. Thanks to Devin Caughey and Hillel Soifer for very helpful comments on previous versions.

* UF Research Foundation Professor & Associate Professor of Political Science, Box 117325 Gainesville, FL 32611, bbsmith@ufl.edu.

> *"Despite the ready availability of these powerful, yet different, modes of inference, historical scientists have often been beguiled by the stereotype of direct experimental proof, and have wallowed in a curious kind of self-hate in trying to ape, where not appropriate, supposedly universal procedures of the scientific method." –Gould, 1986.*

Introduction

It is safe to say that political science has taken an experimental turn. From numerous voter experiments in both American and comparative politics to the establishment of the *Journal of Experimental Political Science*,[1] to two recent books on natural experiments (Dunning 2012; Diamond & Robinson 2010) the construction of and analysis of experiments is likely to be with us for the indefinite future. The arguably greater leverage on causal identification provided by experimental settings, for one, affords inferential confidence. Moreover, for some big questions, constructing actual experiments is neither feasible nor normatively tenable: the outbreak of major episodes of war or conflict, for example. Where field or survey experiments are not feasible, scholars sometimes look for natural experiments, (see for example Dunning 2012, chapter 8; Robinson et al 2009). However, the long-range and macro-causal dynamics of some such episodes in history render statistical analysis difficult or impossible. Below I term such examples natural historical experiments (NHEs). Most of the tools social scientists have for classifying them, and for analyzing them, come from the statistician's toolbox and often fit uneasily onto macro-history if at all. Moreover, there remains some confusion over where NHEs end and the comparative method begins (for example Diamond & Robinson 2010).

My main goal in this essay is to systematize the ways in which comparative historical analysis (CHA) can systematize and improve inference in NHE research designs. First, I draw on examples of island- and border-based NHE research to illustrate that there

[1] journals.cambridge.org/xps.

are multiple types of NHEs beyond the paradigmatic potential outcomes template. These include combinatory (configurative) and critical juncture-trajectory (or path-dependent) variants, each of which can be fruitfully analyzed with comparative historical tools. Moreover, the macro-focus of many CHA methods lends the approach productively to cluster-level (region, district, state, ethnic group, etc.) treatments that are best analyzed at those macro-levels. Finally, NHE research, like natural experiment studies employing large-N analysis (LNA), usually have limited external validity and here I suggest strategies for building multi-method research designs incorporating the approach.

Useful attention has been given elsewhere to the value of qualitative inquiry in validating the as-if random requirement of the natural experiment (Dunning 2012, 212-219) and I seek to build on it here. As such, in addition to validating experiments themselves, I show how they can be better systematized by conscious deployment of CHA tools. Longer-range or natural *historical* experiments are analytically close to the critical juncture framework employed in comparative historical analysis (hereafter CHA) with the main difference being emphasis on agency at the juncture or moment of treatment. Here I build on the work of historical scientists and on CHA to show how the analysis of natural experiments can benefit from explicit use of these methods.

In the next section I delineate the differences between natural historical experiments and standard natural experiments. I define NHEs as distinct from the broader set of natural experiments by virtue of methodological imperatives and by the logic with which they are identified and then analyzed. In this first section I also outline the ways in which scholars otherwise inclined to regression analysis frequently use an under-systematized form of CHA to validate as-if-randomness. Then, I highlight the frequency with which large-N approaches to natural experiment research focus on the wrong levels of analysis, usually to increase the sample size. CHA enables us to focus on the appropriate levels and to build inferences from both causal sequence and purposive case selection and analysis. As a result, CHA may often be better suited to elucidating causal processes or trajectories linking treatments to long-range outcomes. The strengths of CHA for analyzing causal effects at the macro- or clustered level, for parsing out the relative causal import of multiple or bundled treatments at that same macro-level, and for systematizing lengthy historical-causal processes all promise a great deal in the study of large-scale social processes. To the extent that many natural experiments of history are of precisely this type, we stand to gain much by incorporating CHA into natural experiment research.

In the second section I use examples of a commonly employed NHE form in economics and geography—island colonization and its long-range development outcomes—to highlight two alternative forms of natural experiments to potential outcomes models: a) combinatory treatment and critical antecedent effects, and b) trajectory (or path-dependent). I show how the synthesis of comparative historical and experimental research logics lends itself well to understanding and analyzing NHE empirics. I build on the role of CHA in as-if random validation to demonstrate its broad value at each stage of NHE research, showing how it best deals with clustered treatment analysis, the exploration of critical juncture effects and subsequent trajectories or causal pathways. In this section I use the early 20th century division of greater Kurdistan to illustrate these points in sequence. Finally, the fourth section addresses the frequent external validity problems found with natural experiment research, illustrating the common problems faced by all types of natural experiment research designs. In this section I suggest some fruitful multi-method sequences for incorporating NHEs into broader research projects.

NATURAL HISTORICAL EXPERIMENTS IN PRACTICE

As a Subset of Natural Experiments

I performed a Google Scholar search that claimed to turn up 1,650,000 publication results since 2004 for "natural experiments."[2] In practice the appearance of that exact phrase went about 50 hits deep. Of those 50, five (and those five were among the most cited) were specific to political science, with another dozen or so in economics. The range of what scholars mean by the phrase varies substantially. Dunning (2012, 2-3) defines it as follows: "social and political processes, or clever research-design innovations, create situations that approximate true experiments… Causes are randomly, or as good as randomly, assigned among some set of units."

This random, or as-if random, assignment of treatment to some units and withholding from others is the core presumption for confirming natural experiment status. Accordingly, the role that qualitative methods can play in validating the as-if randomness takes pride of place in Dunning's discussion of qualitative research. Indeed, his more demanding practical assessments of the validity of the as-if random claim along a continuum (250), of model credibility (280) and treatment relevance (303) suggests two things. First, it is not just qualitatively as-if random assessed studies that come up short: John Snow's famous (and as qualitative as quantitative) cholera exploration wins the as-if random validity tournament with a pair of quantitative studies coming last. Rather, we see a range of studies (among them Card and Krueger 1994, Grofman, Griffin and Berry 1995 and Posner 2004) that simply do not give this aspect of the research design close attention. Second, more systematic attention to comparative historical exploration of treatment

[2] The direct URL is http://scholar.google.com/scholar?q=natural+experiments&hl=en&as_sdt=1%2C38&as_ylo=2004&as_yhi=2014. Accessed June 26, 2014.

assignment—of the kinds I detail below—could potentially alleviate many of the concerns that render promising natural experiments less viable once turned into research projects.

Table 1 presents the validation strategies for an array of studies.

Table 1 about here

Posner's study of cross-border divergence in ethnic politics among the Chewa and Tumbuka of Zambia and Malawi is, among the studies along Dunning's continuum, the one truly historical in nature—as a result of the long time span between treatment and observed outcomes. To my mind, this may be a part of why it is "ranked" less compelling in each of the above criteria: the temporal distance between cause and effect is such that it is uneasily squeezed into a short-run template of the sort that would accommodate an immediate policy-outcome episode, for example. It is natural experiments like the one leveraged by Posner that suggest another parallel, this time between NHEs and comparative historical analysis.

Validating the "Experiment": What is Different about Natural Historical Experiments

Following broad analytic imperatives summarized by Dunning (2012, 236) as the a) assessment of treatment units/actors' information about treatment conditions, b) incentives to self-select, and c) capacities to self-select into treatment groups, two major empirical strategies for validating the as-if random nature of an intervention's assignment(s)—a major litmus test for any kind of experiment—are common in political science and political economy. One follows the logic of balance tests and explores the covariates of treatment and control groups prior to treatment assignment. In the assessment of as-if-randomness in regression discontinuity (RD) designs, "assignment to the treatment is determined, either completely or partly, by the value of a predictor (the covariate X_i) being on either side of a fixed threshold" (Imbens and Lemieux 2007, 2-3). The presumption of X_i being smoothly, if at all associated with outcome variation is central to finding a causal effect on one side or the other of the threshold. This broad difference of means logic is common in NE research focused on relatively recent treatments and/or on those that naturally produce a large number of observations at the appropriate level of analysis.[3]

However, a second and equally common mode of validation is historical and archival. It relies on plumbing the historical record to confirm that there were no systematic connections between important post-intervention outcome predictors and units' assignment to treatment groups. Although the historical dredge work is often implicit on conditions, the best of it focuses in on the two in my definition of NHEs: lack of knowledge on the part of assigners and no causally relevant relationship between the attributes of units and treatment assignment. Those in charge of assignment (i.e. those with the power to create treatments in the first place) must be considered as well, but less explicitly than might always be appropriate. Related to this is that many policy interventions (treatments) are quite explicitly aimed at producing predictable events and announced in advance, which could allow subjects to self-select. As McCauley and Posner (2015, 411) note, many contemporary events used as natural experiments are methodologically suspect because the treatments are "strategic political decision[s] made in light of [their] anticipated effects."

David Freedman described the kind of contextual knowledge and qualitative detail necessary for analysis of such episodes to succeed as "shoe leather" (Freedman 1991). Dunning refers to qualitative data as "nuggets of information." In general, we see rather less systematic attention to prescription for organizing these data to validate as-if-randomness than for quantitative data. Those of us committed to CHA as a mode of making

[3] Again, large-N samples are not always "natural" or appropriate given the analytics of the question of treatment. Dunning, observes in Posner (2004) that the individual-level survey data analysis assumes unit independence across individuals in Malawi and Zambia while the theoretical logic of Posner's (and Miguel's, above fn.) arguments postulate causal effects at a clustered, or group, level.

causal inference will of course typically expect to have more than this in play: indeed, what distinguishes deep knowledge from CHA is the higher-order deployment of rich historical data across multiple units with careful attention to method-specific rules of procedure as discussed below. In short, organizing qualitative historical data in line with long-standing CHA strategies enables a more robust level of confidence in our inferences, whether they are causal or validating as-if-random treatment assignment.

Natural historical experiments are arguably of a different kind than proximate policy treatment episodes since those applying treatments are often largely ignorant of the long-term consequences of their actions as well as of the attributes of the units to which they are applying treatment. Indeed, this must be the case for an historical episode to fall within the scope conditions of an NHE. Figure 1 lays out how NHEs, and in particular their macro-variants, differ from other comparative frameworks.

Since in political science and economics the demarcation of borders is so often the site of natural experiment research, it is the family of phenomena on which I focus here. At root, in validating the experimental nature of a border assignment episode in history, we ask, “Who made the border assignment, what did they know about the people on different sides, what (if anything) did the people inheriting these borders have to say about them, and how much latitude did they have to move to the other side post-assignment?” Given the typically smaller numbers of clustered units—and the Kurdish minorities I address below numbered four—it is arguably best to begin from a maximally skeptical position. That is, if we begin by presuming the borders were non-random with reference to treatment and control groups, and proceed by looking to confirm that, only in the absence of such historical evidence and preferably evidence that prior attributes of those groups differed in no way related to treatment. This strategy enables us to avoid some intuitive but problematic shorthand rules.

For example, McCauley and Posner, focused on African borders but offering lessons to those of us who work on other regions, suggest as a rule of thumb that "Where borders follow meridians, parallels, or other rectilinear or curved lines, they can plausibly be taken as exogenous to the characteristics of people and places on the ground" (2015, 411). Even here, unless we had such massive numbers of borders as to require a simplifying assumption, we would likely run into unnecessary problems. The Saudi-Iraqi border, running straight and slightly downward of parallel from west to east, is one such problematic case for a straight line rule. This border, whose negotiation in the Treaty of Muhammarah (May 1922) and clarification in the Protocol of Uqayr (December 1922) followed the First World War, also followed years of movement by Arab tribes on both sides, al-Saud military conquests of southern Iraq, and conscious efforts by the British (who ruled Iraq) and Saudi leadership to lay the groundwork in advance for post-border politics, was plainly endogenous to rulers, ruled and conditions on the ground.[4]

Whatever the difficulties in validating as-if randomness historically, it is the case that such episodes are a different kind than the natural experiments whose treatment and effect stages run in tight temporal periods. What is more important than whether we validate them historically or statistically is the long-range spread between treatment and effect, suggesting that exploring them through another, more historically sensitive lens might be more appropriate.

NHEs As a Subset of CHA

NHEs are also a subset of CHA, and within this population comprise a set of designs in which nature rather than scholar establishes the range of cases of study. Posner's (2004)

---

[4] The treaty is available at https://treaties.un.org/doc/Publication/UNTS/Volume%201750/II-183.pdf. Accessed June 24, 2015. See also U.S. Department of State 1971.

exploration of the effect of group size assignment at independence on later ethnic identity salience suggests explicitly that there are long-range impacts that accrue from whether an ethnic group becomes a larger or smaller community within a new state. Miguel (2004) suggests that divergence in nation building across the Kenya-Tanzania border since the 1950s accounts for the differential quality of public goods provision through the mechanism of interethnic cooperation. Berger's (2009) analysis of public goods provision north and south of an administrative line in colonial Nigeria demonstrates a persistent divergence 50 years after independence and more than a century after the division. In a separate setting (India) Banerjee and Iyer (2005) explore the effects of different kinds of British colonial tax systems on late 20th century public goods provision. Darden (n.d.) shows that mid-19th century imperial demarcation continued to shape anti-Soviet insurgency in Ukraine in the 1940s through its effect on national identity across several generations. And, as I explore below, the persistence of Kurdish nationalism in Iraq and Turkey today—and its failure in Iran and Syria—stems in large part from a set of rebellions that began in the 1920s. Two of the most important questions confronting political scientists today—the sources of effective governance and of persistent civil conflict—appear to have deep historical roots. Such long-range causal relationships require not just the validation of as-if random treatment assignment but also a clear theoretical logic to elucidate how original causes produce and reproduce themselves over decades.

Because the intrinsic structure of NHEs involves the nature- or history-driven formation of structured comparisons, they are logically similar to quasi-experimental qualitative research designs. Indeed, to some extent the structured comparison is a methodological effort to approximate the naturally occurring version, just as regression analysis of observational data is a substitute for randomized clinical trials. The logic of the two is compatible, but NHEs are fundamentally different than structured comparisons in the sense that history, rather than the scholar, creates the comparison sets. More accurately, we can think about NHEs as the set of situations or research designs occupying the overlap space between natural experiments and comparative historical case sets (see figure 1). Where they differ is that in NHEs "nature" is applying some or all of both the independent variable (treatment) and control elements in the comparison.

What is often intrinsic to NHEs is a potentially confounding issue: the fact that "big" treatments to one imperial or state political sphere bring with them a multitude of effects that can be difficult to adjudicate—the "what is the treatment?" question. This may be more or less challenging depending on a) extant theoretical context to guide the researcher and b) the extent to which the cases chosen by nature provide a solid platform for adjudicating the causally central treatment factors. When the question at hand is well researched, and if the cases lend themselves to addressing a number of alternative arguments, the analytic logic of CHA can serve a valuable role. Subsequently, closer to the recommendations I develop below is the following take on synthesizing experimental and analytic logics:

> The randomness with which nature intervenes in real-world circumstances may be both harder to ascertain and less perfect than the manipulations of scientists working in a laboratory… a more liberal approach recognizes the inferential advantages of even imperfect natural experiments as well as social scientists' long experience in thinking about inference in terms of ceteris paribus. In other words, we prefer *giving the analyst room to correct some (smallish) level of observed nonequivalence* in treatment and control rather than focusing too narrowly on random selection. (Robinson, McNulty and Krasno 2009, 343, emphasis added).

In the spirit of this methodological imperative, I suggest the following definition: *natural historical experiments (NHEs) are situations in which substantively important treatments are assigned to units in ways that are often as if random in terms of both a) information possessed by the actors who assigned the treatments and b) the prior attributes of the units. The treatment effects are both frequently bundled and typically unfold over long stretches of time with causal dynamics accruing along systematic trajectories. NHEs differ* *from other natural experiments in that the units are usually macro-level, states, regions, ethnic groups etc. or other large "clusters."* I employ and build upon this definition in the pages that follow. Natural historical experiments, then, should arguably occupy a central focus for both comparative historical scholars and for students of natural experiments in the social and political worlds. What remains is to outline a concrete set of principles for analyzing the latter with the tools of the former, and I turn to those in the next section.

COMPARATIVE HISTORICAL ANALYSIS OF NATURAL EXPERIMENTS

Managing the issues that hinder the effective analysis of natural experiments could be facilitated by approaching these empirical phenomena ecumenically. More precisely, we stand to gain by withholding judgment on the choice of methods until we have ascertained the answers to three questions. First, which type of natural experiment is it? As I suggest below, the nearly universal "potential outcomes" template laid over most such historical episodes is but one of three distinct types. Second, what exactly is the theorized treatment? Answering this question under some conditions is better accomplished via CHA than by large-N analysis. Third, at what level of analysis does the theory contend that the treatment took place? While many examples of natural experiment research present LNA data and results at the individual level (for example Posner 2004), the core theoretical insight is often directed at cluster-level effects, in Posner's own framework at the level of ethnic groups within different states. In this section I develop concrete research strategies, drawn from workhorse CHA tools, that can provide answers to each of these questions.
In this section I outline several distinct and sequential research strategies for analyzing natural experiments of history with tools from the comparative historical toolbox. First, I outline some procedures specific to the approach for assessing the validity of as-if randomness. Second, I show how classic CHA can help even statistically driven NE

researchers to adjudicate between bundled treatments to pinpoint which are causally central. Third, I distinguish three discrete types of NHEs, showing how each implores a different set of CHA methods.

Historical Validation of As-If Randomness

An important, perhaps the most important, element of the NHE research design is the moment of treatment. For the NHE researcher the aim is to demonstrate that at a particular juncture in time an historical-political treatment was applied as if at random to some units and withheld from others. With NHE research, however, the questions also relate closely to the reasons for assignment of some units to treatment and others not. The goal is typically to establish that nothing about the treatment assignment (location in one new national state rather than another, for example) is causally related to prior attributes of the units. Frequently this is accomplished by illustrating historically that none of the actors central to treatment assignment decisions possessed the information to assign systematically, or else made simple but random decisions (such as choosing a latitude line to divide jurisdictions or otherwise dividing them in ways unreflective of sociopolitical dynamics on the ground).

Dunning (2012, 243-244) suggests in particular: How much information did they possess about the implications of border decisions? How much capacity did they have to influence the assignment of borders? Finally, how relevant were their incentives to assign in terms of what we now know to have come later? McCauley and Posner (2014, 411-412 at 411) draw on a wide array of African border-based studies to make the following observations:

> Where African borders follow meridians, parallels, or other rectilinear or curved lines (as they do in about 80 percent of cases), they can plausibly be taken as exogenous to the characteristics of people and places on the ground (Englebert,

> Tarango and Carter 2002; Alesina, Easterly and Matuszeski 2011). However, where they follow geographical features such as rivers and watersheds, or where the negotiations that led to their drawing explicitly took account of historical or demographic factors, it is possible that pre-partition settlement patterns caused the populations lying on either side to differ in material ways.

Close historical attention to key actors, to demographic realities and especially to the possibilities of treatment subjects (members of communities on either side of prospective borders) moving back and forth, in short, can help to decide on the central criterion for natural experiment status, whether historical or otherwise.

Interestingly, this strategy is about as common for quantitative NHE research as it is for qualitative studies. Banerjee and Iyer (2005), for example, explore the long-term effects of land revenue systems under British rule in India on post-independence investment in agricultural productivity, health and education in 166 districts. They build the case for the natural experimental nature of this assignment, however, in a way that blends historical research with the comparative method (1195-99). Among other contextual factors they address British knowledge (or misreading) of local elites, the individual influence of British officials elsewhere with strong ideas and little India expertise, prior existence of landlord classes, and finally post-independence factors that could confound colonial era treatment effects. Berger (2009, 6-9), too, uses historiographic accounts to explore and to set aside potentially confounding factors: prior indigenous and colonial institutions, British commercial interests, agricultural profiles, and ethnic demographics above and below the 7º 10’ line. These two examples are strong models of how to preface quantitative treatment effect analysis with rigorous qualitative assessment of as-if randomness: at once contextually knowledgeable and thorough while keeping theoretical interests front and center to guide the exploration of specific alternative explanations.

Miguel (2004, 333) proceeds in a similar but less detailed fashion, simply citing a single source (McEwan 1971) to argue that "the current border was drawn in the period 1886-90 by British and German colonial authories largely ignorant of the ethnic and political entitities that existed in the region." fashion, building the case for comparing districts in Kenya and Tanzania historically, then analyzing quantitative data from those districts.[5] This is not to say that Miguel's claim is questionable, simply to note that without a more detailed explanation of the myriad plausible differences generating the divergence in the two districts of focus we cannot be as confident of the as-if randomness of assignment to either Kenya or Tanzania. Because this mode of validation is so common across multiple methodological approaches, it is all the more incumbent upon us to systematize how we go about it. Dunning's (2012, 243-44) imperatives to explore the "information, incentives and capacities" of actors is perhaps the best starting point. What I suggest we need to add to it is to ask about these attributes for two sets of actors: those who decide on and administer the treatment, and those who are treated. The former, in the case of border delineation, are going to include whomever decides on the borders (colonial officials or imperial diplomats) and then the political elites who rule the divided territories (imperial rulers, different administrators and/or rulers of newly independent states). The latter will be the subjects or citizens of the newly divided territories: in some cases they will have agency in the delineation, and in others their agency will be limited to the capacity to self-assign to one side of the border or the other. The partition of India, and the mass of migration (voluntary and forced) that accompanied it, serve as one highly visible example of a border assignment

[5] Note, however, that since the treatment was clustered at the national level this disaggregation to districts within Kenya and Tanzania (in search of greater numbers) faces a level-of-analysis problem similar to the one Dunning observes in Posner (2004).

that, although somewhat random in terms of boundary location, was far from it in terms of treatment and control populations.

More often than not these situations result from elite actors imposing sets of conditions (treatments) on non-elites, putting power center stage and thus squarely in line with the affinity of macro-historical scholarship for power relations. However, in classic critical juncture analysis, CHA scholars tend to privilege the agency of central actors to the "story" about to unfold, showing subsequently how the result of that agency was to catalyze path dependent processes. NHE scholars, especially those who look at post-colonial or post-war border assignment, take initial actors for the most part as information-challenged treatment assigners (nature's unwitting Principal Investigators) whose jurisdictional decisions catalyze equally long-term political dynamics for others who lack agency. Sometimes treatments are applied to units in great enough numbers to analyze statistically and other times not, and sometimes the appropriate units of analysis are large enough to result in relatively few. The absolute number, however, is less important than the long-term causal trajectories or paths that take shape. As a result, establishing the as-if randomness of an NHE must often take a different form. Here, given the clear methodological proximity to structured comparison and to CHA, it is appropriate to invoke research strategies from the latter.

In the next several paragraphs, I employ empirics from a broader research project to illustrate some approaches to, and some obstacles to, using CHA to validate as-if randomness. That broader project asks why some ethnic minorities manage to sustain rebellions against central governments for decades, while other similarly situated groups fail to. One part of it is an in-depth look at Kurdistan from its post-World War I division to present. As the First World War approached its end, British and French officials began to ponder how to maintain their respective nations' influence in a post-Ottoman Middle East.

One part of the waning Ottoman empire—what surrounding Arab, Persian and Turkish empires and the Kurds themselves had come to imagine as a relatively coherent greater Kurdistan—became the site of what I am going to argue below is an illuminating natural experiment of history. With little attention to cultural, geographical or prior political fault lines in Kurdistan, the French and British, and later to a lesser degree nascent Turks, created four new Kurdish ethnic minorities in four new states ruled by Persians, Turks or Arabs.

Two key sets of actors—the British, French and Turks on one hand, and different sets of Kurdish actors, elite and non-elite on the other—must be assessed along different sets of systematic criterial questions. In the case of Kurdistan, the British and French, as victorious powers with a decided power advantage over the nascent Turkish political elite, made the decisive choices that resulted in the borders in place today. But, between 1920 and 1923 the new Turkish government gained voice as well. We know that the British generally possessed more information about Kurdistan than either the French or Turks. For example, one oil-rich part of what is today Iraqi Kurdistan—the Mosul region—is where it is because the British wished to assign that part of Kurdistan to their own mandate territory rather than to the French. But other than that we have little reason to believe that the current borders are related in any non-random way to the attributes of different parts of Kurdistan at the time.[6]

[6] This border creation took place over three separate agreements: The Sykes-Picot accord of 1916, the Treaty of Sevres (1918) and the Treaty of Lausanne (1923), the last of which was a dramatic change from either of the first two accords. The new leaders of the Republic of Turkey participated only in this last agreement and, while Kurdish representatives had been involved in the run-up to Sevres they were excluded from Lausanne. The Ottoman-Persian border, ultimately separating Iran from Iraq and Turkey, had been agreed upon by 1914. See Ateş 2013, Arfa 1966 pp. 26-32, O'Ballance 1973 chapter I, and McDowall 2004 pp. 117-146.

What about the Kurdish notables who strove to secure for themselves and Kurds in general the best possible outcome they could? Here we know, again from the rich historical record that at least in the case of oil that the British hid this from Kurdish elites and that they were all but left out of the negotiations that ultimately led to these borders. In short, we have a strong as-if random case to make for the border assignment strictly in terms of information by those who did the assigning. I turn to broader sets of questions about the four Kurdish regions below.

Adjudicating Between Bundled Treatments

Whether the by-nature treatment lends well to structured comparison can be assessed by using the logic of purposive case selection as a check on the appropriateness of the case set. More precisely, CHA allows the scholar to correct nonequivalence through the careful and purposive analysis of a set of cases analytically nested in extant theory, which allows for the identification of the factors in a bundled treatment set most likely to be causally central. NHEs are structurally similar to critical junctures in the sense that treatments for some units narrow the range of options open to them, or change the range of options relative to other units. Also shared across the NHE-CHA divide is a predilection for variants on Mill's method of difference research designs. After all, as-if random treatment is only valid once assigned to one group and withheld from an otherwise similar control group, just as CHA scholars choose cases in order to hold as many other possible explanations constant as they can. This is especially useful for NHEs in which assignment—to one country instead of another for example—brings with it a bundle of treatments. An NHE that took the form of a critical juncture in time might arise around a single unit—for example an ethnic community—then divided between two jurisdictions.

Assignment to one state rather than another, of course, carries with it a potentially large number of causal factors for adducing long-range outcomes.

Following partition, the Kurdish minorities (ethnic "clusters") in Iran, Iraq, Syria and Turkey were "treated" to bundles of different factors in subsequent years, many of which scholars of ethnic violence find to be important determinants of separatist rebellions. I have presented them in tabular form in Table 2. At a structural level, all four Kurdish regions are mountainous, a covariate not in line with subsequent divergence. Two regions—Syria and Iraq—are rich in oil, a pair again not co-varying with the presence of long-term rebellions. Two Kurdish minorities—in Turkey and Iran—were ruled by center-right ethno-dominant autocracies that sought to minimize sub-identities. The other two—in Iraq and Syria—were ruled by Ba'athist Arab socialist regimes that sought to cultivate the rhetoric of cross-communal fraternity while in practice suppressing ethnic minority autonomy. Finally, in subsequent decades Kurdish organizations in two of these states—Iran and Iraq—received periodic great power support. This two does not predict the variation in outcomes which we seek to explain.

The point of outlining these bundled treatments across four Kurdish minorities is to highlight the value of comparative historical analysis for overcoming what could otherwise be a daunting degrees of freedom problem. So what *does* co-vary with the outcome—that is, why is it that Kurds in Iraq and Turkey have fared so much better over time than their counterparts in Iran and Syria? Three of the four new Kurdish regions (Iran, Iraq, and Turkey) were characterized by strong intra-ethnic social hierarchy at the moment of national assignment. As a result, Kurds in these three countries were endowed with traditional or "start-up" social structures that could facilitate mass mobilization. That would prove crucial as new rulers in Iran (Reza Shah), Iraq (King Faisal) and Turkey (the Kemalists) began to engage in precocious Weberian state building, seeking to exercise a political monopoly over their territories before they had achieved a legitimate monopoly of violence.

Kurdish elites claiming the authority to speak on behalf of their communities in all three countries rebelled in the 1920s against their respective central states, and all three were defeated. However, only in Iraq and Turkey were the Kurdish defeats followed by a concerted state effort to reform agriculture in rural Kurdistan. In these two states that led to subsequent urban migration over the next two to three decades. In Iran, no such large wave of migration took place, leaving most of the Kurdish population still in rural areas by 1960. When serious state-led development began in all three countries in the 1960s, it catalyzed the emergence of new urban social movements across the political spectrum, including Kurdish ones. The relative lack of an "urban Kurdistan" in Iran, however, meant that a small insular and now more politically radical elite had no such city-based constituency as its counterparts in the Kurdish-dominant cities of Turkey and Iraq.

What this meant is that in Iraq and Turkey, Kurds by the late 1960s and early 1970s were a majority in a number of cities, building new social networks and undergoing the common process of political identification, "crystallization," and radicalization that we see so much else elsewhere (see for example McAdam 1982; McSweeney and Jokisch 2007, 171-172, 2015, 23-26). This, the broader project argues, explains the emergence of such vibrant resistance by the Kurdish Workers' Party in Turkey and the Patriotic Union of Kurdistan in Iraq in the 1970s. Note here that the unfolding of historical-causal process over several decades would not fit, nor it would be amenable to productive analysis, by potential-outcomes or comparative statics. Rather, only by taking a consciously trajectory-driven approach to comparing post-treatment dynamics is it possible to understand how initial treatments—at an almost unarguably *critical* juncture—produced the processes that drove Kurdish nationalism from elite to mass volume in just two of these Middle East states.

The partition of Kurdistan following the First World War affords us the opportunity to explore how careful use of CHA can isolate the key causal factors in an expansive bundle of treatments. When British, French and ultimately Turkish negotiators settled on the post-war borders of Iraq, Syria and the remnants of Ottoman Turkey, they did so with limited knowledge of the contours of Kurdish society and politics. As a result, the "Kurds" found themselves by the mid-1920s living in Iran, Iraq, Syria or Turkey. Kurds in all four of these new states had by 1945 created Kurdish Democratic Party organizations—following the original model of the KDP-Iraq created by the Barzani family in the 1920s. Kurds in Iraq, Iran and Turkey staged rebellions against their central governments between 1924 and 1936 and, in the most ambitious formal claim to date, Iran's KDP declared an independent republic in Mahabad in 1946. Yet, by the early 21st century Iraq's and Turkey's Kurdish minorities, spoken for by new organizations with substantially different goals, were by far the most successful in pressing political claims. Explaining this long-range divergence begins with an effort to understand just how each Kurdish minority began and under what conditions.

Many of the most consistent determinants of civil conflict today do not accord with the cross-Kurdish political experience. In Iraq, Iran, Syria and Turkey Kurdish minorities inherited, unbeknownst to them, differential resource endowments (the Kurdish regions of Iraq and Syria are oil-rich; those in Iran and turkey are not). All four Kurdish regions are mountainous, a key determinant of civil war onset (Fearon and Laitin 2003). In pairs they were also "treated" to two systematic regime and nation building projects: Ba'ath Arab socialism in Iraq and Syria and center-right ethnic homogenizing efforts in Iran and Turkey under Reza Shah and Kemal Ataturk. A more fine-grained exploration of state policies toward the four Kurdish minorities reveals little practical difference below these regime macro-types, too. Arab, Persian and Turkish rulers all stifled meaningful cultural expression and autonomy for their Kurdish citizens, often resorting to severe repression, and during the 20th century Kurds in each of the four states were excluded from access to representation at the national level (Cederman et al 2010; Hassanpour 1992). In short, many of the usual suspect factors we would expect to help explain the divergence fail to show much promise as central treatment effects.

Turning to factors internal the new Kurdish minority groups in each country, three of the four (in Iraq, Iran and Turkey) were characterized by strong hierarchical rural social structures.[7] In those three states there were tribal or elite-led early rebellions against efforts at state penetration, which were followed by state military responses. The socioeconomic changes catalyzed by these early conflicts set in motion yet again differential trajectories of state-minority conflict. Table 2 systematizes these factors. I turn in greater detail to the trajectories below, but here I want to focus on how close attention to each treatment in a sizeable bundle can be addressed and, in many cases, set aside for lack of fit with outcome. In essence, we take a natural experiment—and both the as-if randomness of border drawing and little subsequent migration across those borders provide strong plausibility evidence here—and lay the structured comparison and critical juncture-trajectory frameworks over it. In doing so we are able to put to one side all but the treatment factors that turn out to be causally central, at least in the cases under exploration. In this case most of the plausible factors that were part of the treatment

[7] Syria's Kurdish minority was forcibly displaced there during the interwar years, moving into territory where it had no long-standing presence. As a result, there existed no established elite class capable of using hierarchical mobilizing structures to organize an effective mass response to Syria's rulers at independence.

applied to Kurdish minority groups fail to predict successfully the divergence of outcomes, while one set does track with them.

Table 2 about here

There are, as illustrated in this comparison, some temporal analytic issues as well. Observing that three of the four new Kurdish minorities rebelled, and that only two managed to sustain viable transitions from elite to mass nationalist movements (in Iraq and Turkey), begs the question why Iranian Kurds failed to sustain a mass movement subsequent to the Mahabad declaration. In the next section I turn to a specific discussion of the contribution of critical juncture and trajectory approaches to elaborate.

Natural Historical Experiments as Critical Junctures, Treatment Effects as Trajectories

In close temporal natural experiments treatments are relatively simple. A minimum wage change in one American state and none in a similar bordering state, for example, can arguably produce effects in well under a year. Such episodes fit appropriately into a potential outcomes template. In many natural historical experiments, by contrast, both the bundling of treatments and the possibility of interactive causation suggest that there are other considerations to take into account when analyzing treatment moments. One is that the historical kinds that often interest political scientists do not always fit the single-treatment potential outcomes template, in which we have otherwise similar units subjected to a one-time uniform treatment. Instead, as I detail below, natural experiments follow three systematic patterns, two of which are arguably better analyzed through CHA. One of these two is combinatory causation, in which a treatment may be necessary but causally sufficient only in combination with prior unit attributes, or what Slater and Simmons (2010) term critical antecedents. The other is duration or trajectory causation. In a duration case it is not treatment per se (such as colonial rule) producing effects but the span of time during which treatment is in effect. In a trajectory case, treatments unfold in causally deep ways over long time periods, such that only by focusing on causal depth rather than correlative breadth can we illuminate the ultimate effects.

Whether they focus on agency or on critical determinants at a moment in time, scholars of critical junctures share in common a focus on particular moments of change. And, in the same way that potential outcomes natural experiment research usually hinges on the reactions of individuals to treatments, some critical junctures focus on the uncommon expansion of agency that emerges. On the other hand, not all junctures or treatment moments are so contingent on agency and are instead more structurally driven. These are the same kinds of critical junctures in which proximate causal factors interact with prior unit attributes to produce outcomes jointly. We may think of these, rather than being like lab experiments involving people, as being analogous to a chemistry experiment in which one substance (treatment) is added to a number of subject substances and in which the treatment effect is jointly produced.

Slater and Simmons (2010, 889) refer to the prior unit attributes as *critical antecedents*: "factors or conditions preceding a critical juncture that combine with causal forces during a critical juncture to produce long-term divergence in outcomes." Framed in this analytic lens, treatment effects may interact with, and activate, latent causal efficacy in units' attributes, so that the treatment operates in part indirectly. With reference to the case of post-partition Kurdistan, the critical antecedents were a) intra-ethnic hierarchy and b) the presence of major urban centers in each Kurdish region. When each of the four were treated with homogenizing state and nation building projects, the effects were combinatory: only those Kurdish minorities possessing both strong mobilizing structures and in-region cities embarked on state-minority trajectories that produced long-term success of nationalist movements. Figure 3 outlines both the combinatory treatment effects and trajectories that resulted.

[Figure 3 about here]

Whereas political scientists often look to borders as sites for natural experiments, geographers and economists have regularly analyzed islands. These borders—ocean shores—are truly natural.[8] Yet, as analysis of one such study reveals, they share much analytically in common with border-based NHEs. Rolett and Diamond (2004) begin with the puzzle of Easter Island: why was this place abandoned? They situate Easter Island in the full population of islands colonized by Polynesians, asking why some were a) more prone to deforestation and agricultural collapse and b) why some continue to thrive today while others like Easter are uninhabited.

This natural experiment leverages the as-if random exploration by Polynesians of islands throughout the South Pacific—discovering new islands and attempting to settle them. While it begs the question whether the treatment is a) random treatment application of Polynesian colonizers to dozens of islands, or b) treatment application of different island conditions to dozens of sets of Polynesian settlers, the effect either way is a combinatory one. The units of analysis are therefore occupied Polynesian island colonies—human-natural ecologies—and their proneness to deforestation the outcome whose divergence Rolett and Diamond seek to explain. Controlling for environmental determinants of plant growth—rainfall, latitude, ash dust drift from Asia (tephra), and terrain age—they find that shoreline reef strongly shaped deforestation levels. Makatea, an "uplifted reef formation of sharp, fissured coral" is both unfriendly to plant life and dangerous to walk on.

[8] Chinese island building in the South China Sea notwithstanding.

As a result, islands rimmed by makatea in part or in whole were less likely to be settled, and thus less likely to be deforested.

Because it was the crucial interaction of Polynesians and land, “Easter’s collapse was not because its people were especially improvident but because they faced one of the Pacific’s most fragile environments” (Rolett and Diamond 2004, 445). In short, makatea mattered greatly but only when combined with an island’s encounter with Polynesian colonizers. The structural-environmental factors in this analysis are thus alternative explanatory or control variables, but also ones that operate independently of human presence, highlighting the combinatory human-ecological nature of the makatea effect. Figure 4 models this causal relationship. Diamond’s and Rolett’s work on greater Polynesia illustrates not a classic “treatment” but rather a uniform treatment—Polynesian settlement—applied to islands with very different *critical antecedents*. Here it was the physical and environmental attributes of different islands that determined what the (combinatory) effect of Polynesian discovery would be. The treatment differentials are thus configurative ones and display the same kind of attention to interactive causality that also characterizes much critical juncture scholarship.

[Figure 4 about here]

Colonial Treatment Duration and the Prosperity of Island Nations

Thus far we have explored critical juncture, trajectory, and configurative or combinatory dynamics in natural historical experiments. A final variant, again fruitfully analyzed with tools from comparative historical analysis, relates to duration effects and asks not whether or not but for how long. Feyrer and Sacerdote (2009) ask: why are some post-colonial island nations around the world so much better off than others? Employing a similar research design to Rolett’s and Diamond’s—albeit one that takes in all of the world's island nations rather than simply Polynesian ones—Feyrer and Sacerdote suggest that during the Age of Sail, when wind was the sole source of energy for seafaring imperial expeditions, natural wind patterns assigned European discovery treatment to some indigenous island populations while sparing others until the advent of steam engines. The historical experiment here was truly natural.

The authors find that islands colonized earlier tended to be so as a result of as-if random location along prevailing wind routes. Moreover, they are more likely to be richer in the present.[9] The first task here is to validate the as-if random nature of wind pattern location and then to ask whether it is related to the onset of colonization. It is. Once that is established, the authors turn then to exploring possible mechanisms that could provide a causal link between early colonization and later prosperity. The elucidation of these mechanisms is an application of comparative historical logic to quantitative data analysis.[10] It is sensitive to the less valorous side of what might be driving the early colonization effect—the near or complete extermination in many island colonies of indigenous populations and domination of settlers. Feyrer and Sacerdote also take care to outline the fundamental differences in the goals, and therefore the strategies, of earlier (extractive commercial) versus later (transformative comprehensive) waves of colonization. In short, this study provides a set of valuable, portable research steps for building NHE research designs.

First, the authors make a strong case for as-if random assignment and tie the treatment/control groups compellingly to important long-range outcomes. While wind is a kind of ideal type natural force for treatment assignment, it is only causally important here

[9] At the risk of what Slater and Simmons termed "infinite regress," it is reasonable per Diamond to ask whether prior to colonization some islands were already better endowed by nature, then simply better positioned to thrive during and after colonial rule.
[10] Feyrer and Sacerdote 2009, 246-48.

to the extent that it can be shown both to relate closely to the timing of colonization and to those long-range development outcomes. Second, they situate the NHE treatment in its comparative historical context, taking the simple question of colonization timing and embedding it in a larger set of questions related to the content of colonization. Finally, as the authors move from one set of cases to another, and from one level of comparative analysis to others, the implications of each set are then used in the next. Feyrer and Sacerdote are centrally concerned not with the presence or absence of treatment but rather its *duration* as the causally crucial variable. What mattered in their analysis was not what kind of colonialism reigned on any given island but simply how long, again providing an instructive link to the temporal concerns central to CHA research (see for example Aminzade 1992).[11]

NATURAL HISTORICAL EXPERIMENTS AND MULTI-METHOD RESEARCH DESIGN

The major weakness of most natural experiment research in political science is external validity. The same as-if or truly random treatment assignments that give us greater confidence in the causal relationships at work also hamstring our confidence in the portability of theoretical conclusions. Posner's (2004) analysis of cross-border variations in the political salience of Chewa and Tumbuka ethnic identities in Malawi and Zambia might imply a greater comparative role for group size in ethnic mobilization, or it may not. Berger's compelling identification of a long-range public goods differential on either side of a randomly chosen latitudinal line in colonial Nigeria might successfully predict similar divergence elsewhere, or it might not. The divergence of today's four Kurdish minorities'

[11] To reiterate, they do not contend that the content of colonialism was irrelevant; just that when accounting for it the effect of colonial *duration* continues to be strongest.

fortunes in the Middle East might portend a broader comparative role for early rebellions and resultant trajectories: like these other examples, however, we cannot hold much confidence in their external validity without taking these insights further afield. Incorporating them into multi-method research designs holds promise to do just that, and to synthesize the unique contributions of all three forms of NHEs with the strengths of other approaches.

Recognizing which kind(s) of NHE are at work or manifest in a particular setting will inform how it may be best integrated into a multi-method project. We are not necessarily analyzing a single NHE with multiple methods, but also plausibly using comparative historical or other methods to analyze the NHE itself while using additional strategies to explore additional empirics. While I deal primarily here with NHEs that are analyzed using CHA, the same basic guidelines would apply if we were employing regression-driven research and then moving to more in-depth cases to trace causal processes or to systematize trajectories.

In the last section I introduced the combinatory treatment effect NHE, in which a treatment effect is causally important only in interaction with antecedent unit traits. In the cases of post-partition Kurdistan, I showed how homogenizing nation-state projects (the treatments) catalyzed long-range opposition to themselves only against Kurdish minorities with a) established social hierarchies and b) in-region urban centers. I also showed how one environmental feature of Pacific islands—makatea, the sharp reef shoreline—shaped ecological outcomes only in interaction with Polynesian settlers (the treatment uniform to the islands). A number of productive options exist for multi-method research elaboration with such cases. One—and a commonly used one in econometric political research—is simply to build a set of statistical indicators for each component and then interact them. In the case of the theoretical argument used to explore Kurdish nationalism across Iraq, Iran, Syria and Turkey, I used demographic ethnic indicators from the Ethnic Power Relations Project to capture urbanization and social structure data from the updated Ethnographic Atlas project for hierarchy. These became central indicators in a dataset incorporating about 450 ethnic groups in Africa, Asia and the Middle East that also draws on “deep historical conflict” data back to 1400. Together the historical-causal insights afforded by the NHE that was Kurdistan’s partition enable a much broader analysis of ethnic demographics, early conflicts and contemporary mobilization. More abstractly, this approach to testing the argument more broadly takes the key concepts from an initial CHA analysis and broadens the scope (while reducing the causal depth). It is similar to the methodological progression in Lieberman (2003) and to Smith (2007), with the key difference of course being that in the latter two studies scholars rather than nature determined the selection of cases for close study.

Another possible, and probably more challenging, approach would be to seek out an equivalent NHE to study with further CHA. Stripping the case of Kurdistan’s division down to what it “is a case of”—here the division of a demonstrably coherent ethnic region into parts of multiple states in a way that is potentially (not necessarily) as-if random—might yield additional cases. Balochistan—divided into parts of Afghanistan, Iran and Pakistan in the early 20th century—is one such possible case, as is the Touareg region of Africa’s Sahel. However, one almost uniform cautionary in scholarship on natural experiments is that they are rare. That fact begs the question: must additional cases for CHA be potential natural experiments? The answer is no. If the question is why some ethnic minorities sustain long challenges while others fail to, and if the theory at hand points to clear causal factors, then simply analyzing other ethnic minority groups could be just as fruitful.

The other major type of NHE addressed above is the trajectory form, in which initial treatments generate not just initial differences but systematic long-range historical pathways. Since the arguments in such examples are validation-dependent on demonstrating that such trajectories take shape when initial conditions evince, testing for external validity requires similarly exploring additional cases to assess the extent to which those trajectories do appear and unfold similarly. The key question in testing the breadth of the trajectory is whether ethnic minority groups' capacity to sustain long-range rebellions is a function of a) early ones catalyzed from without by state builders and enabled from within by mobilizing social structures and b) the same sequential patterns of contentious mobilization over time. An alternative or compliment would be to use a trajectory assignment model with multiple subsequent junctures/change moments. Here we could envision building the same project with an initial, starting treatment moment and a carefully theorized and explored set of processes and trajectories rather than a single treatment moment and outcome moment.

Recognizing the appeal of not just causal depth but breadth as well, though, one can opt to extract some "snapshot" indicators from key historical moments in a theorized trajectory. Drawing on recent innovations in both the production of and use of data to explore the "deep historical roots" of conflict (i.e. Fearon and Laitin 2014, Dincecco et al 2014), we might for example explore whether there are decades-long connections between late 19th or early 20th century conflicts involving ethnic groups as a function of their social structures, then ask in a second-stage analysis whether these factors help to explain conflict proneness in the late 20th and early 21st centuries.

CONCLUSION

Perhaps due to early natural experimental method work by quantitatively oriented scholars, or perhaps due to skepticism that natural experiments are more than a "chimerical template," comparative historical researchers have lagged in their deployment of this potentially powerful research design. This essay has endeavored first to show how useful CHA can be not just for establishing the as-if randomness of treatment assignment but at multiple steps in analysis. Given the large number of econometric studies that rely entirely, though implicitly, on CHA techniques for validating their NHEs, this is a natural area for beginning to integrate the two. Second, it has outlined how analytically close many natural experiments are to some common approaches to CHA. The logical affinity of treatment moments and critical junctures, for example, helps to illuminate some important ways in which CHA may be used to elucidate exactly why some treatment moments are so crucial for subsequent trajectories. Moreover, the three distinct types of natural historical experiments discussed here—focusing on combinatory treatment moments, trajectories, durations and antecedent conditions, respectively—parallel the development of CHA as a methodological enterprise. Third, the partition of Kurdistan as an NHE has shown how valuable the overlay of CHA over such an historical episode can be for separating out the truly important treatment factors from less important ones when many are bundled as happens following national border assignments. Finally, I have outlined here some ideas for embedding NHE research into a broader multi-method design, showing how its strengths can complement both qualitative and quantitative observational methods and highlight long-range causal processes.

On this final point, it is important to note what I am not up to here, which is to suggest that CHA can or ought to try and trump quantitative NHE research designs. Quite the opposite: each approach is going to carry with it a set of laudable strengths and inevitable limitations. But, where CHA is uniquely well suited to the tasks at certain points in the analysis of natural experiments of history, those strengths hold the promise of capitalizing on these rare moments in time and making the analytic most of them. In addition, the systematization and tracing of causal mechanisms and processes (qualitative analysis) vs. statistical regularity (quantitative analysis) offers a number of promising avenues for making NHE analyses part of a broader multi-method project. Single treatment models are at times too comparative static-dependent for historical analysis. As McAuley and Posner nicely pose the challenge: “the problem with experiments is that while they may establish the existence of a causal relationship, they leave unresolved the knotty issue of explaining why it exists.” McCauley and Posner (2007, 13) At the same time, the turn to deep historical origins-focused research in political science and political economy suggests that incorporating insights from multiple methodological perspectives has much potential. Natural experiments of history deserve an important place in that research agenda.

Figure 1.

	Historical/Macro	Non-historical/Macro
Natural Experiment	Natural Historical Experiment	Standard Natural Experiment*
Non-natural experiment	Standard CHA	Standard Comparison**

*these include Dunning's "standard" natural experiments but also regression discontinuity and instrumental variables NEs. **These include structured comparisons as well as regression analysis of observational data

Table 1. Mode of validation for Natural Historical Experiments

Author	NHE Type	Validation of NHE method	Level of analysis	Dominant Method
Banerjee & Iyer (2005)	Land tenure system assignment	(CJ) Historical record: colonial context-ignorance	Sub-state districts	Regression analysis
Berger (2009)	Tax system assignment	(CJ) Historical record: random latitude assignment	Households; local government areas	Regression analysis
Darden n.d.	Imperial division in Ukraine	(CJ) historical record	Ukrainian provinces	
Diamond (2010)	Border division on Hispaniola	(CJ) historical record	Countries	Narrative
Feyrer & Sacerdote (2006)	Timing & Duration of Colonization	(CJ) Instrumental variables (wind); historical record	Small island nations	Regression analysis
Miguel (2004)	National border assignment in Kenya and Tanzania	(CJ) Historical record: colonial context-ignorance	Districts	Regression analysis
Smith (2010)	Imperial collapse & border division in Kurdistan	(CJ) historical record colonial context-ignorance	Ethnic minorities	Comparative historical; integrated comparative

Figure 2. Natural Historical Experiments as Subsets of Both Natural Experiments and Structured Comparisons

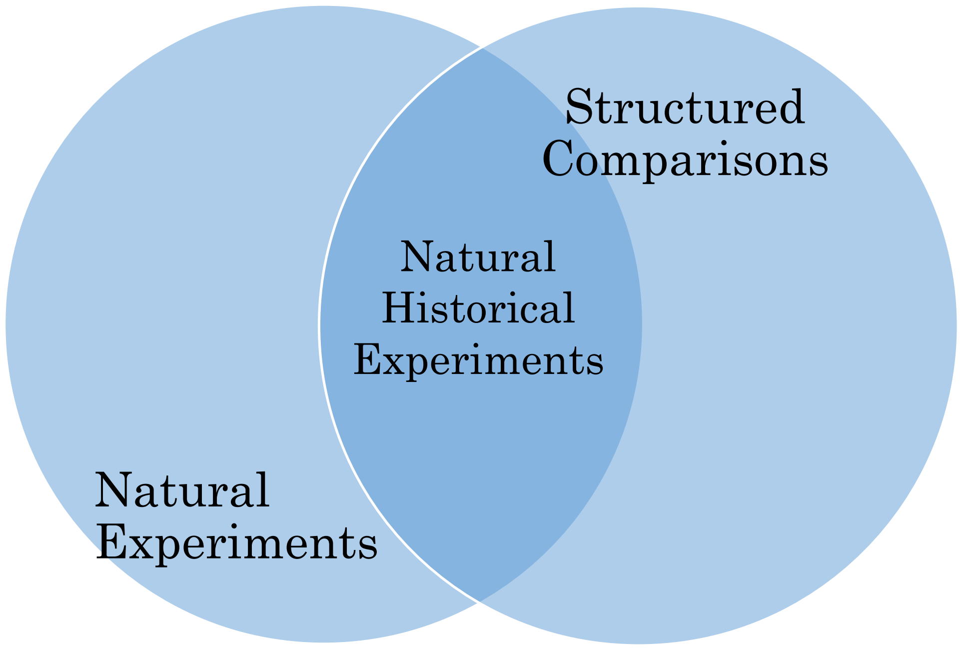


Table 2. Clustering & Bundling of Historical Causality:
Evidence from the Partition of Kurdistan

	Iran	Iraq	Syria	Turkey
Background Conditions				
Ottoman legacy?	0	1	1	1
Persian imperial legacy?	1	0	0	0
Resource rich region?*	0	1	1	0
Mountainous?*	1	1	1	1
Critical Antecedents				
Strict intra-ethnic hierarchy at moment of treatment?	1	1	0	1
Large Kurdish-majority Cities?	0	1	0	1
State Building Treatments	1	1	1	1
Rival Explanations				
Ba'ath Arab Socialist regime	0	1	1	0
Ethnic Nationalist regime	1	0	0	1
First-stage outcomes				
Early post-independence rebellion?	1	1	0	1
State Repression & Forced Rural Restructuring?	0	1	1	1
Urban migration	0	1	1	1
Long-range outcomes				
Late 20^{th}/early 21^{st} century rebellions?	0	1	0	1

Factors that predict rebellious trajectories are shaded. * baseline model variables either for civil war (Dixon 200X) or ethnic rebellion (Cedermann et al 2010)

Figure 3. Combinatory Treatment Effects on Trajectories in Kurdistan

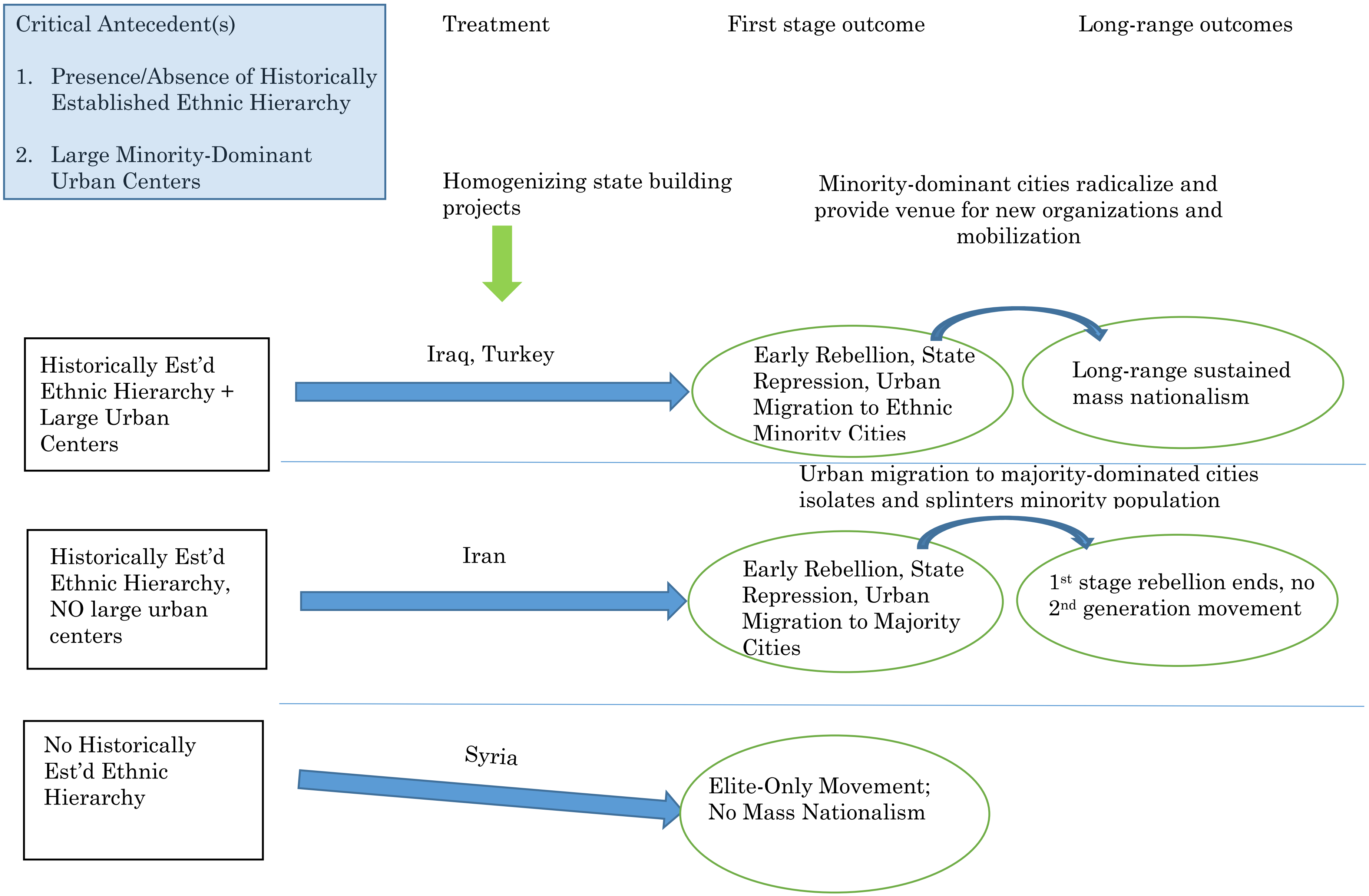
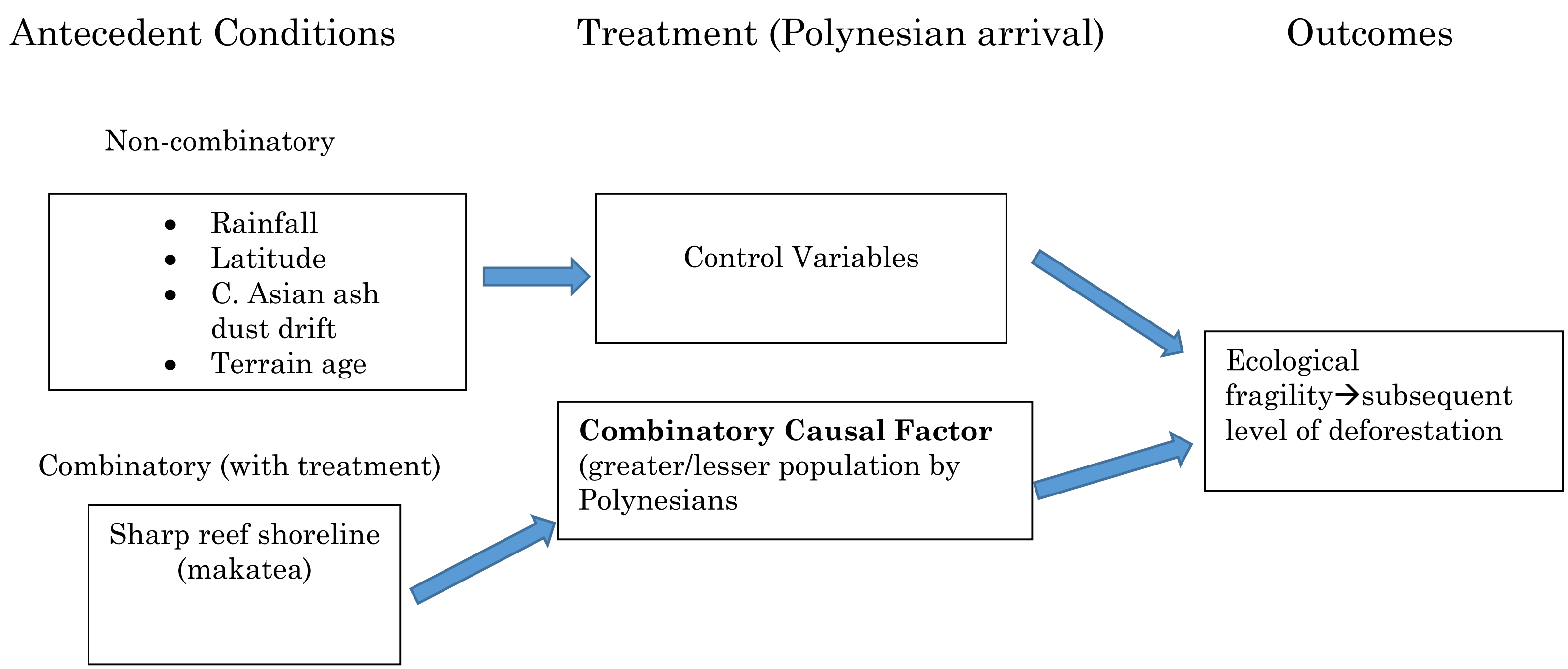


Figure 4. Combinatory Treatment Effects and the Easter Island Puzzle.



References

Ateş, Sabri. 2013. *The Ottoman-Iranian Borderlands: Making a Boundary, 1843-1914.* New York: Cambridge University Press.

Banerjee, Abhijit and Lakshmi Iyer. 2005. "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India." *The American Economic Review* 95, 4 (September): 1190-1213.

Berger, Daniel. 2009. "Taxes, Institutions and Local Governance: Evidence from a Natural Experiment in Colonial Nigeria." Unpublished manuscript, University of Essex.

Capoccia, Giovanni and R. Daniel Kelemen. 2007. "The Study of Critical Junctures: Theory, Narrative and Counterfactuals in Historical Institutionalism," *World Politics* 59, 3: 341-69.

Darden, Keith. N.d. "Resisting Occupation: Lessons from a Natural Experiment in Carpathian Ukraine." Unpublished manuscript, American University.

Diamond, Jared and James A. Robinson eds. 2010. *Natural Experiments of History*. Cambridge, MA: Belknap Press.

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

Feyrer, James and Bruce Sacerdote. 2006. "Colonialism and Modern Income – Islands as Natural Experiments." *NBER Working Paper Series* No. 12546. http://www.nber.org/papers/w12546

Hassanpour, Amir. 1992. *Nationalism and Language in Kurdistan, 1918-1985*. San Francisco: Mellen Research University Press.

Hyde, Susan D. 2007. "The Observer Effect in International Politics: Evidence from a Natural Experiment." 60, 1 (October): 37-63.

McCauley, John F. and Daniel N. Posner. 2015. "African Borders as Sources of Natural Experiments: Promise and Pitfalls," *Political Science Research and Methods* 3, 2: 409–418.

Meyer, Bruce D. 1995. "Natural and Quasi-Experiments in Economics." *Journal of Business & Economic Statistics* 13, 2 (April): 151-61.

Miguel, Edward. 2004. "Tribe or Nation? Nation Building and Public Goods in Kenya versus Tanzania." *World Politics* 56, 3 (April): 327-62.

Robinson, Gregory, John E. McNulty, and Jonathan S. Krasno. 2009. "Observing the Counterfactual? The Search for Political Experiments in Nature." *Political Analysis* 17, 4: 341-57.

U.S. Department of State. 1971. Iraq-Saudi Arabia Boundary. *International Border Study* no. 111. Washington, D.C: Bureau of Intelligence and Research.