Theory, History, and Political Economy

Sean Gailmard*

Professor, Department of Political Science, University of California, Berkeley, 210 Social Science Building, Berkeley, CA 94720, USA; gailmard@berkeley.edu

---

ABSTRACT

In this short paper I offer conceptualizations of history, theory, and their interplay under the aegis of political economy. My primary argument is that historical political economy (HPE) depends on theory for its success. First, notwithstanding empirical causal *identification*, causal *explanation* is impossible without theory. Second, establishing *causal uniqueness* (that a particular mechanism is the only reasonable candidate to explain a case) and *causal generalization* (that a particular mechanism explains other cases not yet empirically analyzed) are exclusively theoretical in nature. No research design, however rigorous or credible, can support these claims; thus it is counterproductive for HPE scholars to analyze cases simply because causal identification is possible. I conclude that political economy can best contribute to historical understanding by applying mechanisms from this field as candidate explanations of important cases, and history can contribute to political economy by helping us discover new mechanisms.

Launching the first journal specifically devoted to a field of inquiry is an important part of defining what the field is. That is what the *Journal of Historical Political Economy* is doing. For historical political economy, the field, to succeed we need to understand what it is.

It does not help that political economy itself is subject to multiple interpretations. The meaning is relatively clear in the economics discipline: political

---

---

economy is the application of economic methodology to questions about politics, whether or not economic factors are part of the story. Economists who describe themselves as doing political economy typically mean this. In political science the meaning is broader. When a political scientist says that they study political economy, they might mean roughly what an economist would mean. Or they might mean that they study the interplay of politics and the economy: the political determination of economic structures, or the effect of economic conditions on political outcomes. This latter usage is more about questions than methods, which run the (wide) gamut of those used throughout the discipline. When a sociologist uses the label, he typically means the latter interpretation of political scientists.

Adding the modifier "historical" to political economy compounds the ambiguity. Historical political economy can mean the use of historical data (whatever that is) to evaluate theories in political economy, in the same way and for the same purpose (whatever that is) that one would use "contemporary" data. Historical political economy can also mean the application of political economy to understand historical events and processes in and of themselves.

When we do historical political economy (HPE), we must remain open to all these self-definitions. Social science will never and should never have a master methodology. Ambiguity of meaning may not be desirable, but it is reality. Accepting reality is better than artificially imposing a doctrinaire definition for the sake of pretend clarity.

Nevertheless, these disparate interpretations of historical political economy do share a common attribute. In all of them we are trying to understand the causal relations among actions and events in history. This commonality in turn reveals the primacy of theory in two ways. First, causal explanation requires theory. Identifying a causal effect from some randomization, discontinuity, and the like somewhere in history says nothing in itself about the reason for the causal effect. Only theory can explain why it happens. Second, generalization of causal mechanisms beyond cases where they are identified also derives solely from theory. The credibility of causal identification within a case — whether from randomization, discontinuity, process tracing, narrative, etc. — has nothing to say about it.

In this paper I explain these positions. First I address the need for theory in causal explanation, and review the meaning of "historical" and "theory" as I use them. Next I argue that all theory-oriented empirical work in HPE answers essentially one question: whether specific causal mechanisms are or are not consistent with observed case evidence. Furthermore, I consider whether empirical work can ever justify statements of *causal uniqueness* (that a specific causal mechanism is the unique best explanation of case evidence) and *causal generality* (that a causal mechanism explaining one case can also explain others). I answer negatively and contend that such statements are exclusively theoretical. I conclude that HPE scholars should select important

cases and try to understand the range of causal explanations consistent with them, not select cases on the basis of identifiability of some causal effect.

The Need for Theory in Historical Political Economy

For the identification-oriented empiricist, it may seem tempting to skip all the verbiage to follow. After all, theory is not necessary to identify causation in the sense of a treatment effect under the potential outcomes model of causation. Causal identification comes from manipulation, not theoretical mechanisms.[1] Science is replete with cases where a treatment's causal efficacy on an outcome was established long before any mechanism was credibly articulated.[2] In this grand tradition, many a paper in subsequent pages of this soon-to-be-august journal will surely consist of exploiting a discontinuity or randomization in treatment assignment found in history to identify a causal effect of something on something else.

The question will always be why this deservers our attention. It takes grit and shoe leather to demonstrate treatment effects with any credibility, but treatment effects are not unusual in the world. Things have causes and there are lots of things, so there are lots of causes. The reason they deserve our attention is that they matter.

Treatment effects matter when they help us *explain* things or when they *generalize* to a broader class of situations. Explanation of empirical findings and generalization from them are both impossible without theory.

Scientific explanation involves answering "why." This implies causation: a cause and effect relationship between a factor and an outcome (cf. Ashworth *et al.*, 2021; Elster *et al.*, 1998, p. 9; Gerring, 2006, p. 5; King *et al.*, 1994, p. 75; Van Evera, 1997, p. 9).[3] However, demonstrating causation does not in itself explain *why* a causal relation exists; that is, what causal mechanisms

---

[1]The identifying assumption in observational causal inference research is always a theoretical statement, but this statement rules out *other* factors to explain an effect of one variable on another. It does not identify a mechanism by which a treatment affects an outcome.

[2]Three examples: (*i*) The health benefits of hand washing, which Viennese physician Ignaz Semmelweis noted some 20 years before Pasteur's breakthrough on the germ theory of disease. (*ii*) The life-saving benefit of insulin for Type I diabetics was verified experimentally decades before the exact action by which the body regulates its entry into cells, or what it does inside a cell, was established. (*iii*) The analgesic effects of aspirin were identified experimentally, and the drug was successfully sold commercially, some 80 years before the mechanism of action was identified. These discoveries surely all followed from some theoretical conjecture, but not anything that could be called a mechanism.

[3]It is also standard in empirical social science to take the potential outcomes model as the definition of causation. I have no reason to disagree with this, but also no commitment to it versus other theories of causation. In any case, the definition of causal mechanism subsumes a definition of causation, so none must be provided here.

do (Little, 1991, p. 178; cf. Clarke and Primo, 2012, p. 153).[4] Treatment effects are not mechanisms, they are the observable result of mechanisms. A treatment effect, like any empirical association, is a descriptive statement about the world that itself requires explanation. Explanation requires a mechanism, not just evidence that a mechanism operates (Little, 1991).

A major reason why credible causal inference matters is because it demonstrates causes of a broader class of events that are important. Consider the field-defining work of Acemoglu *et al.* (2001), a classic of HPE. While this paper has shined bright light on New World colonial institutions, its intent is to identify random variation in these institutions for the purpose of identifying random variation in *contemporary* political institutions, and then identifying the effect of these institutions on contemporary economic conditions (cf. Acemoglu *et al.*, 2001, Figure 1). The importance of this paper is not limited to causal identification in the cases in their data. It is that these cases are held to be representative of a broader category of political institutions, and that these institutions are held to affect economic performance in general. As I argue below, there is no way to draw a connection between the cases in the data and the broader class of interest without theory.

Another aspect of the need for theory in HPE is a question of comparative advantage. When social scientists study historical events in and of themselves, we must be cognizant that a (probably vast) literature in historiography also studies those events. It would be irresponsible, both to our history colleagues who have done this work and to the spirit of intellectual curiosity, not to seek out that literature and understand what it says. What can we add to the conversation that our colleagues in history do not?

A significant comparative advantage of political economy in the study of history is theory — or rather, a specific approach to theory. Historical narrative generally embeds a theory as social scientists would understand it (a claim I defend below), but historians and social scientists have different tools and perspectives for constructing theories. Theoretical political economy is essentially a "library of mechanisms" (Guala, 2005).[5] In the game theoretic approach that my own work draws on, this library includes canonical elements such as collective action problems, coordination problems, commitment problems, signaling and information transmission, information aggregation, preference aggregation, strategic voting, electoral competition, agenda setting, coalition building, political agency problems and the like. Other social science approaches draw on different mechanisms — critical junctures, path dependence, policy feedback, network structures, etc. — but the process is the same. Faced with a new situation that requires explanation, PE scholars

---

[4]Mechanisms also simultaneously give a constructive answer of *how* causation happens.
[5]Or, as Cox (2004) has it, a "cookbook of standard models."

deploy and combine elements from this library to make sense of it. Historians do not typically have the same background in adapting and combining these mechanisms to specific situations, so PE scholars can do so to contribute to historical understanding. When the combination of mechanisms deployed is qualitatively unlike others already in our library, we also have a contribution to theory in PE. Thus, the exchange between theoretical PE and history is bi-directional.

What Makes HPE Historical?

The necessity of theory in historical political economy does not establish the meaning of either "theory" or "historical." I consider the "historical" part first and turn to "theory" next.

Alas, historians and epistemologists have not managed to come up with a definitive characterization of "history," and I do not aspire to fill this lacuna. Nevertheless, it is worth briefly reviewing the issues in characterizing what we are up to.

Perhaps the simplest view of "historical" is that it involves events that occurred in the past. This is not a particularly useful view to distinguish HPE from PE, or history from anything else. On this view, all data are historical. To be observed and analyzed, events must have taken place at a time before the analysis takes place. Another name for "a time before" is "the past."

A slight revision makes a bit more sense: "historical" applies to events that occurred in the past under temporally bounded social or institutional configurations no longer in operation in the place where the event occurred. Applying this conception to 19th century American politics makes some sense of the difference between "historical data" and "data from the past." Analyzing American governance under the "state of courts and parties" is historical because it points to a specific institutional configuration that structured behavior but no longer exists, and the point of the analysis may be how we transitioned from that institutional configuration to a subsequent one. On the other hand, estimating DW-Nominate scores from early Congresses assumes the same process governing votes, under the same institutional structure, that one would use to estimate DW-Nominate scores from the most recent Congress. This is not "historical" so much as it is an analysis of data from long ago.

So, perhaps "historical" means both "in the past" and "from a process different from the current one in some salient respect." Yet there are several issues with this view. First, it *assumes* a relevant social or institutional process no longer operational, but part of the point of historical research (whatever it means) may be to *discover* whether that is true. If we found

that in fact a process underlying some event from long ago worked exactly the way contemporary processes do, we would not then revise the research to be "non-historical." Second, this conception is unproductively constraining. Analysis of State Department archives from the 1990s may be "historical" even if the processes behind the events and decisions reflected in them are very much still operational.

Both of these views involve temporality of the data with respect to the present. But they may still assume non-temporal processes within the past. For instance, suppose one characterized the "state of courts and parties" as a more or less static regime where things happened in a particular way (different from how they happen now). On this view "the state of courts and parties" would timeless within itself, even if it existed before our current regime. This treats the past in a stilted manner.

To resolve this we can take "historical" to involve temporality *within* a process, regardless of how long ago it occurred. This means that history involves events that unfold over time, in sequence, with awareness at later points that something might have happened at earlier points. The focus may be on change in social or political events over time, or on explaining the absence of change over time; the "over time" part is key. On this view, research does not become "historical" simply by virtue of relating to events long ago. Historical requires taking seriously the possibility of change over time within a case.

For historians, this may be the most satisfying abstract characterization. For HPE, it is a bit constraining too. If one provided a comparative analysis of political institutions in ancient Greek city-states as a function of their internal resource endowments and external security position, we would probably consider it "historical political economy" even if all variables were treated as entirely static.

Rather than nomological characterizations of history, one can take a "historical realist" interpretation. History is what historians do, and historical data is data generated by the research practices of historians (Mandlebaum, 1938). In practice, this involves archives and the triangulation of multiple sources. It also involves sensitivity to the inferences implied by the existence of data: this usually means someone wanted it to exist, and it may be possible to infer something about the underlying social and political relations in question from this fact. More importantly, this interpretation requires understanding that historians bring their own often implicit theoretical commitments to the table, rather than producing theory-free bodies of facts to be explained (Lustick, 1996).

All of these views have some merit except for the simple first one. In the absence of a sharp characterization, my view is that HPE should take a relatively broad view of what falls within the field.

What Is Theory in Political Economy?

To accomplish the goals of this paper, it is necessary to establish what I mean by "theory." At the outset I note from casual ethnography that some scholars have strong feelings about the distinction and relationship between the words "theory" and "model." I do not. I accept that phrases such as "rational choice theory," "game theory," and the like connote general approaches to social inquiry. Theory in this sense has nothing to say about anything; one needs a model to make the approach concrete (e.g. in game theory, specifying a game form, utilities, actors, and a solution concept). Yet given that we develop models and call them theories, there is little at stake in referring to a specific model in (e.g.) game theory as a theory in itself, rather than a model. If one insisted on a distinction and hierarchy we could just as well call models "theories" and game theory a "paradigm" (though not in a Kuhnian sense). In any case, in what follows I refer to "theory" generally in this latter sense, and it is effectively interchangeable with "model."[6]

In this paper's usage, a theory explains a causal relationships between stated factors and one or more outcomes, by reference to a causal process. In contemporary social science, such a causal process is often referred to as a *mechanism*. The causal process or mechanism is held to explain how the factors cause the outcomes. In this usage, the factors correspond to the *explanans* of a theory, the outcome to the *explanandum*, and a mechanism represents the social process linking them (Little, 1991).

Not all statements of regular relations between variables are theories according to my usage. Theories entail causal relationships, and an explanation for that causation. A description (or a prediction) of a relationship between two variables in a dataset is not a theory, because it does not specify whether one causes the other or they share a common cause, and it does not specify a causal process. A prediction that a decision maker will make a particular decision is not a theory, because it does not explain the reasons why — the mechanism. A treatment effect identified with all the precision social science can muster is not a theory because it does not self-explain. My usage of theory is catholic but not vacuous.

Even restricting attention to causal statements, it is possible to establish *causation* without causal *explanation*. In an empirical tradition still embraced by many social scientists, causation is empirically constituted by regular relationships among observables under appropriately controlled conditions. The demands of establishing causation in this sense are rigorous — we must achieve all-else-equal comparisons — but surmountable. Experiments, regression discontinuities, and the like aim for this.

---

[6]Cf. Dowding (2015).

Important as knowledge of causation is, it does not explain *why* X causes Y (Shepsle, 2009). It does not explain the cause of the effect, the reason why the effect exists. By positing a causal mechanism, a theory, and only a theory, provides a reason why a treatment effect comes about. It provides a conceptually compelling reason for an empirical, causal link between X and Y. This position is sometimes called "causal realism," reflecting the idea that causal forces, distinct from regular relations between variables, are real things. In the language of directed acyclic graphs (DAGs), variables are nodes and may be related by causal mechanisms. Mechanisms are the reasons for directed links between nodes, but not attributes of a DAG itself.

This focus on mechanisms draws on an increasingly common perspective over the last several decades in the philosophy of social science (Elster *et al.*, 1998; Goertz, 2017; Hedström and Swedberg, 1998). In turn these scholars follow philosophers of science like Nancy Cartwright, Wesley Salmon, and others (Cartwright, 1999; Machamer *et al.*, 2000; Salmon, 1998; Woodward, 2002), e.g. "Causal processes, causal interactions, and causal laws provide the mechanisms by which the world works; to understand why certain things happen, we need to see how they are produced by these mechanisms" (Salmon, 1984, p. 132); "[T]hings and events have causal capacities: in virtue of the properties they possess, they have the power to bring about other events or states" (Dupré and Cartwright, 1988). On this view a social event is potentially *explained* when a social process is presented that could plausibly cause it. That process is a causal mechanism.

While there is widespread agreement (though not universal) that causal mechanisms are real social forces that explain a relationship between factors X and outcomes Y (Falleti and Lynch, 2009), there is less agreement and more confusion about the ontological status of a mechanism. This confusion leads to conceptual ambiguities about the kinds of inferences that various research designs — statistical analysis, cross-case qualitative analysis, or process tracing — allow us to draw from empirical data.

Many scholars seem, at least implicitly, to conceive of mechanisms as observables or variables. The assumption that mechanisms are observable resides in any research design aiming to measure, identify, or reveal on a causal mechanism. For instance, we hear that causal mediation analysis (Imai *et al.*, 2011) and process tracing (Blatter and Haverland, 2012; Haggard and Kaufman, 2012; McAdam *et al.*, 2008) allow scholars to identify causal mechanisms. This perspective presupposes that causal mechanisms are observable, for we cannot measure or identify what we cannot observe.

Against this argument, scholars who deploy these methods may distinguish what they *observe* (events, actors) from what they *infer* (mechanisms). It might seem that this resuscitates the concept that mechanisms are unobservable. It does not. If one holds that a mechanism can be identified — that a method, at least if successful and well executed, will recover or "isolate" the mechanism

responsible for events — then the distinction is of no consequence: a mechanism is observable. The distinction between observance and inference simply means that some factors are measured with one kind of activity (reading archives, journals, letters, speeches) and the mechanism is measured with another kind of activity (process tracing, mediation analysis). The process trace or mediation model *is* the mechanism-meter: you put in a "case" and it tells you the attribute of the case. That is what any measurement device does.

Conceptually, treating mechanisms as measurable variables essentially turns them into mediating variables in a DAG. There is nothing per se incoherent about treating a causal mechanism as a special kind of observable, but it is inconsistent with the premise of causal realism. If causal explanation requires a mechanism, and a mechanism is observable, then causal explanations consist exclusively of regular relations between observables. We are back to a purely empirical theory of causation, albeit with special new names for some observables. If one actually means the causal realist postulate that causal explanation does not inhere solely in relations between observables, the perspective that mechanisms are observable is no help. To a causal realist, empirical research is useful for constructing possible mechanisms to explain empirical events, or for evaluating the ability of postulated mechanisms to explain those events.

There is also ambiguity about mechanisms among formal theorists. Many scholars agree that models, somehow, capture causal mechanisms (Ashworth *et al.*, 2021; Goertz, 2017; Gordon and Simpson, 2020; Lorentzen *et al.*, 2017). There is less agreement on exactly what a mechanism is inside a model. Paine and Tyson (2020) suggest that formal models embed multiple mechanisms that are parsed out in comparative statics (10). Goemans and Spaniel (2016) contend that a causal mechanism is an entire equilibrium of a model, not just a single comparative static.[7] A third perspective, which I adopt, is that mechanism is a social–behavioral process by which a factor X causes an outcome Y. In Nash equilibrium analysis, coordination of beliefs about strategies is an important, unobserved, and usually unexplained part of the mechanism. Take a factor such as a responder's reservation value π in ultimatum bargaining. A mechanism explains the causal effect of π on responder's share of the pie p. In my view this effect is explained by a conjunction of three elements: (*i*) the game structure (sequence and information sets), (*ii*) the payoffs of each player and their objective to maximize them, and (*iii*) the belief of player 1 that player 2 will act on this objective for any offer.

---

[7]Defining a mechanism as an equilibrium leaves open the question of what a mechanism is if one does not employ equilibrium solution concepts, e.g. rationalizability. If a mechanism is generalized to mean "the solution under the solution concept employed," then one must conclude that in a game with a unique rationalizable outcome, rationalizability and Nash equilibrium are the same mechanism, which does not seem right.

Given these elements, we expect $p = \pi$; changing any of these elements would lead to a different relationship between p and π. Thus, these elements explain how π affects p.

Some of this disagreement may arise simply because the concept of a mechanism itself is subject to disagreement (Woodward, 2002). In addition, in giving a causal account, there is always some elasticity in what is treated as a "cause" and what is treated as a "background condition." For example, one can say the "mechanism" responsible for differences in ultimatum and dictator bargaining allocations is the responder's veto power, while leaving the behavioral and belief coordination postulates in the background.

Theories, Equivalence Classes, and Dimension Reduction

A theory is a statement of the factors that matter in producing outcomes in a category of cases. The theory defines the category of cases, it defines the factors that matter for a particular outcome in those cases, and it specifies the process that connects the factors and outcomes.

The category of cases to which the theory applies is an *equivalence class.* The equivalence class states that cases meeting certain conditions are governed by the theory, and cases not meeting the condition are not. The equivalence class of a theory means, "in cases like A, X causes Y, but not necessarily in cases unlike A." In empirical work, this equivalence class is often identified with "scope conditions": measurable factors that determine whether the logic of a theory is supposed to apply or not.

Defining the equivalence class of a theory is an important part of theorizing. It declares the set of cases to which the theory should be expected to apply. This is crucial for recognizing which cases are "alike" under the theory and which cases are "unalike." In other words, defining typologies of cases is inherently a theoretical statement.

Another important part of theorizing is *dimension reduction.* For any case in the equivalence class of a theory, a participant in or observer of the case could identify innumerable factors or dimensions that could in principle cause the outcome. To be useful, a theory must reduce this set to a usually small number of factors that actually do matter. Without this reduction in factors, a theory is simply claiming that some combination of all factors present in the case caused the outcome. This is not an informative statement, and it does not provide any better understanding of how an event comes about than direct observation of the event unaided by any theory. If direct observation without theory is sufficient to establish the factors responsible for an event and the mechanism that links them, then theory is superfluous. If theory is not superfluous, then direct observation is not sufficient. If direct observation is not sufficient, then theory cannot merely recreate it; it must focus on a subset of factors.

Therefore, a useful theory of an event must be less complex than the social process in which the outcome occurs that the theory purports to capture. "Less complex" means that the theory contains fewer factors as potential causes than one could in principle observe. In this way, a theory enables an observer to understand an event occurring in a social environment with myriad interrelated factors, as actually caused by only a few of them. The reduction in explanatory factors that cause an outcome is part of what it means to understand the outcome.

In the sense that theorizing is a dimension reduction exercise, every theory is necessarily "unrealistic."[8] Dimension reduction means that a theory explains an outcome as a result of a few factors, and there is a loss of realism by excluding all the other factors that an observer of or participant in a process might notice. When a theory fails to include various possible factors as causal determinants of an outcome, the theory is not declaring that those factors do not exist. It is declaring, implicitly or explicitly, that they do not matter. The theory states that, of all the factors that an observer might notice in a situation, it is only necessary to notice certain ones if one wants to understand why an outcome occurs.

Such statements made under a theory might be wrong. A theory might exclude factors that do affect the outcome in an important way. Pointing out the importance of excluded factors is one of the jobs of a critic. The key point is that a critic can never accomplish this task merely by observing that a theory is "unrealistic" in the sense of excluding certain factors from a causal mechanism. Observing this exclusion is equivalent to declaring that a theory has been offered — for all theories come with dimension reduction — and no one would consider this a critique. The critic must also point out that the excluded factors matter for explaining the outcome. That point necessarily entails a different theory. Therefore, one cannot critique a theory without presenting another theory, at least implicitly.[9]

Rational Choice Theory: Rigor, Intelligibility, and Epistemic Humility

Given the constituent fields of political economy, much of the theorizing that occurs in HPE will likely draw from strategic models of rational choice. Everyone can agree that at some level the postulates of these models are unrealistic. An early attempt to systematically apply rational choice theory to historical explanation (Bates *et al.*, 1998) was harshly (and I would argue

---

[8]As Black (1962) put it, models are necessarily "unfaithful" representations of some aspect of the phenomenon being modeled (p. 220).

[9]This is not simply a Lakatosian point that a theory must never be (or is never) rejected unless it is replaced with another. The point is that it is not conceptually possible to argue against a theory without a different theory in mind.

uselessly) criticized for these reasons (Elster, 2000). Why are we still doing
this?[10]

Rational choice in all its variants, like every other theoretical postulate,
is a matter of dimension reduction. Of course the postulates are false at
some level. This tells us nothing in itself about what we can explain with
these postulates: a postulate does not have to be true in all cases to explain
something. Postulates like Bayesian rationality or sophisticated strategic
inference can fail in some settings and yet offer compelling explanations in
others. In any application one can only hope that they are not too inaccurate.
We know that they are when they fail to give any compelling explanation for
events. The same could be said of any other postulate, rational, or otherwise.

Yet this is essentially a negative justification of rational choice. It is the
same justification one could offer for any theoretical postulate. What is the
positive justification for rational choice versus other unrealistic postulates one
could invoke? Sometimes rational choice is justified as especially rigorous, or a
requirement of scientific progress, e.g. "We believe that rational choice offers a
superior approach because it generates propositions that are refutable" (Bates
et al., 2000, p. 700). But rational choice modeling in HPE has no special claim
on analytical rigor or scientific progress. Rational choice might be *compatible*
with scientific progress, but there is no reason to believe that such progress
demands it. Likewise, there are lots of ways to make rigorous arguments and
rational choice scholars do not have a monopoly on them.

In my view there are two important virtues of the strategic–rational choice
approach, and neither is scientific. The first is that it necessarily renders events
intelligible. If we can see decisions of others as the result of goals, constraints,
and strategic incentives as they saw them, then we can understand why they
made those decisions (Collingwood, 1946; Dray, 1957; Weber, 1949). In this
way rational decisions are self-justifying and can never make an observer
incredulous (Satz and Ferejohn, 1994).[11] When they are destructive, we
can lament the environment that produced them, but we understand them
nonetheless. In contrast, if we view the decisions of others as unrelated to
their incentives, or worse as *undermining* their own interests as they saw them,
those decisions provoke incredulity, and explanations that provoke incredulity
are unsatisfying.

The second virtue is that rational choice theory embodies *epistemic humility*.
Rational choice is a commitment to assuming that decision makers understood
their problems at least as well as we, the analysts, do (Myerson, 2006). To

---

[10]The payoff of rational choice theory is distinct from the payoff of formal models. I
argue elsewhere that this payoff comes from assurance of logical consistency and improved
clarity of causal mechanisms (Gailmard, 2020).

[11]This does not confine us to claim that historical actors never made mistakes or miscal-
culated, provided that evaluation is done ex post based on information the actor did not
have.

pronounce a decision as irrational requires us to reconstruct a decision problem and show that, given the information available when the decision was made, the decision undermined the actor's own interests as it understood them. This necessarily means that the decision maker did not understand the problem as well as we do. Since the decision-makers in our cases lived and breathed the events we study (and often risked blood and treasure over them), and since especially in HPE our reconstruction of the decision is based on partial and selected materials, this takes more than a bit of hubris.

These are in some sense humanistic, not scientific, values. If one finds intelligibility and humility important, rational choice is useful. These values are not uniquely satisfied by rational choice theory, but it does satisfy them.

Theory and Data in Historical Political Economy

With some parameters on the meaning of history and theory, the primary question before us is what we can learn by integrating them under the aegis of political economy. I argued above that the theoretical content of political economy is a "library of mechanisms." Given this, the interplay between history and theory in political economy can proceed in two ways. First, we can observe historical events and relationships, and try to reconstruct a theoretical mechanism that produced them. Second, we can build theories of how certain types of processes operate and evaluate them using historical data. Roughly speaking, the first enterprise starts with historical cases and attempts to build a theory to account for them; the second starts with a theory and attempts to evaluate it with historical data.[12]

I will refer to the first approach as "theory-from-data." This approach contributes to theory development by constructing mechanisms to explain case evidence. History as it relates to politics is replete with scenarios in governance, collective choice, preference aggregation, bargaining, and conflict. Any that is not already covered in the positive theory literature can form the basis for a contribution to that literature.

I will refer to the second approach as "theory-to-data." For quantitative approaches, empiricists will see historical data as expanding the frame of reference in which natural experiments, randomizations, and discontinuities can be found to test theories (Cantoni and Yuchtman, 2020; Diamond and Robinson, 2010; Gordon and Simpson, 2020; but see Sekhon and Titiunik, 2012). For qualitative approaches, "The purpose of case studies is to explore

---

[12]These polar cases are ideal types. In practice all theory-from-data research requires some prior theoretical commitment to define "a case" as following a process separable from other "cases." And all theory-to-data research starts with some prior empirical understanding to build a theory. Some research is a more thorough hybrid, e.g. Gennaioli and Voth (2015).

causal mechanisms at the heart of theories" (Goertz, 2017, p. 1), and such cases are the bedrock of narrative history.

Despite obvious differences, I will argue that the order of considering theory and data is irrelevant. Both approaches make essentially the same contribution to historical understanding: they both tell us whether particular causal mechanisms from our library are candidates to explain a particular case or cases. This is an important form of understanding. But social scientists usually want to go further: we want to know if a mechanism is the only or most compelling account of a set of cases (*causal uniqueness*) and whether a mechanism accounts for a wide array of cases (*causal generality*). On the former I will argue that, while not all research designs are equally good at restricting the set of mechanisms, unique causal identification is always an exclusively theoretical statement. On the latter I will argue that non-inductive causal generalization not only requires a theoretical statement, but also a statement that presumes the very causal knowledge that generalization seeks to provide. Consequently, empirical causal knowledge is necessarily local to the set of cases from which it is drawn.

Building a Theory-from-Data

In the theory-from-data approach, the process begins with evidence from a case or cases. The case involves a group of actors, a subset of whom will play an important role in causal explanation. The case involves a sequence of actions over time by these actors, some exogenous events (demographic change, instability in another nation state, etc.), and possibly statistical relationships between events and actions. The researcher chooses the actors, decisions, events, and relationships on which to focus ("stylized facts"). The problem is to explain the observed combination of the outcome variables. Explanation happens by constructing a model with embedded mechanism that accounts for the observed data. Models provide internally consistent interpretations of the data under the constraint of the paradigm from which the model is drawn.

Obviously, there are several downsides to this approach. It pays no attention to cases besides the one to be explained, so may not help for developing theories that can explain many cases. Moreover, this approach is open to the charge — warranted or not — of creating the dreaded just so story. However, there is at least one upside to this approach: by construction, it produces explanations for important events.

Lizzeri and Persico (2004) on English middle class franchise is an HPE classic deploying a theory-from-data approach.[13] They adduce four key facts from 19th century England about the disposition of public spending, intra-elite

---

[13]Other recent examples in HPE include Acharya *et al.* (2018); García Jimeno (2016); Gailmard (2019, 2021).

conflict over public spending, and a shift in the basis from political competition from patronage to proto-party loyalty (p. 712). Then they argue that existing explanations cannot account for all four facts, and develop a new model which can. The model provides a strategically evocative causal mechanism that translates an increase in the value of public goods into voluntary franchise extension, and the paper argues at length that this mechanism is consistent with an important case. It expresses no ambition to explain any other case, but as with all models the logic is applicable to any other case that happens to fall in its equivalence class.[14]

A Stylized Example. Imagine a case in which we learn one actor has a pile of a divisible resource P and proposes to share $p < P$ with a second actor. The case material informs us that the second actor will receive some payoff $\pi < P$ if she rejects the offer. And it further informs us that $p = \pi$. But it tells us nothing else. Our task is to reconstruct a model such that this observation occurs under the specified conditions.

Such a model will postulate a structure of interaction, outcomes from various actions of the actors, objectives, and beliefs, and by this model we explain why $\pi = p$. We observe some facts to discipline the model: actor 1 made a proposal, actor 2 reacted to it, and $p = \pi$. The model must include these facts; everything else is assumed.

One obvious model to explain this case is the subgame perfect equilibrium in ultimatum bargaining: π is common knowledge, p is a take-it-or-leave-it offer from the first actor observed by the second, the players earn $(1 - p, p)$ if the offer is accepted, they earn (π', π) with $\pi' < 1 - \pi$ if the offer is rejected, both players seek to maximize their material payoff, player 1 is aware of player 2's motivation, and player 2 cannot precommit to reject any offer at least as big as π. This model — game structure, objectives, and beliefs — is the mechanism by which the observed facts are translated into the outcome $p = \pi$.

There are two important points about this explanation. First, it adds a lot of structure on top of the observed facts. We need additional structure (though not necessarily *this* structure) for causal explanation. Even if we magically knew that increasing π causes an increase in p (which the case does not tell us, but the postulated mechanism does), that causal information is not self-explaining. Furthermore, even this additional structure may not be

---

[14]Theory-from-data can involve more than one case. Coşgel *et al.* (2012) model the differential response of Ottoman versus European elites to the printing press based on exogenous differences in the relationship between religious and secular authority in each case. Alesina *et al.* (2020) identify general patterns of changing military organization, public policy, and nation building from an aggregate of cases in 19th century Europe, then build a model to connect these observations. Engerman and Sokoloff (2002) develop their theory of endowment determinism in New World colonial institutions to explain variation in structure across those cases.

convincing to everyone: one might reasonably ask how it came to be that player 2 received π in case of bargaining failure, as that is crucial for the outcome.

Second, the observed facts cannot uniquely identify this mechanism as the sole valid explanation, and nor could any other combination of observed facts. For example, the objective to maximize payoffs at every possible information set and beliefs about what player 2 will do at every possible information set are not observable even in principle. Even if the case facts also included player 2's diary stating, "My objective is to maximize my material payoff at every information set," player 2 has not necessarily stated her objective in the game. Instead, she has taken an action in a game between herself and whomever she thought might read the diary. We can assume she never thought anyone would read it and wrote her true objective, but then our assumption — a theoretical postulate — is doing the work of identifying an objective. In the end, we cannot escape that a mechanism is a theoretical entity, not an observable one (George and Bennett, 2005; Goertz, 2017).

More generally, any game containing the observed facts and with $p = \pi$ as an equilibrium outcome is an explanation of this event. I will hazard that there are not many substantively interesting cases where the stipulated facts uniquely identify a game whose equilibrium set contains the observed outcome. The problem of mechanism identification is compounded by elasticity in assumptions about goals (are they always subjective expected utility maximizers?) and belief coordination (can we consider self-confirming equilibria or rationalizability?).

The problem of mechanism identification is infinitely multiplied as soon as we step beyond the narrow confines of game theory. For example, suppose it so happens that $\pi \approx \frac{P}{2}$. Then we would observe that player 1 offered a roughly equal split of the pie. We could just as easily conclude that the explanation is about a norm of sharing or perhaps reciprocity, and the strategic value of the outside option π has nothing to do with it.

With these points in mind, what have we learned about the case? We have learned one candidate mechanism to explain it. Because mechanisms can be applied to many cases, we thereby learn something about how to determine whether other cases are "similar" to this one. We have also learned that the observed facts are consistent with strategic rationality, and therefore that the actions of the individuals in it, as well as the outcome itself, are intelligible to us through this lens. If the explanatory model were new to the modeling literature, we would also have a novel entry in our library of mechanisms.

As an example of historical cases that we actually want to explain, this case is trivial. Just one facet of triviality is that the case included no irrelevant facts; all observed facts were reflected in the explanation. Real cases include an embarrassment of information, which requires judgment about which information to include in a causal mechanism — further compounding

the problem of identifiability.[15] Yet even here we face serious problems of mechanism identification.

Under the HPE umbrella, political scientists, sociologists, and historians have contributed the most to methodological discussion of causal explanation of individual cases. I discuss some of this work below.

Process Tracing. Process tracing is an approach developed in political science and historical sociology for building a causal model to explain a case. Scholars have recently made significant strides in providing rigorous definitions and foundations for this method. Disparate definitions remain, but some key points emerge (Collier, 2011; George and Bennett, 2005; Goertz, 2017; Hall, 2003; Waldner, 2012): process tracing involves a sequence of observations within a single case or a small number of cases. Close observation of the case and the timing of events within it are important. Process tracing can be understood to produce a one-row dataset of concatenated variables defined inductively from studying the case and not commensurate with variables from any other case; thus it cannot even in principle be considered to be one slice of a larger panel dataset. Something in the case can be considered "an outcome." Inference about causation and causal mechanisms is one of the major goals of process tracing. To accomplish this, process tracers do not examine correlation of events in the concatenated dataset; instead they seek to explain the joint occurrence of the concatenated events. Through the within-case focus, process tracing seeks to avoid problems of mis-identifying causal explanations by aggregating cases that in fact are not alike in their causal process. However, process tracing scholars explicitly prioritize the development of theories applicable to other cases. This is a form of causal generalization, which I consider below after discussing theory-to-data analysis.

Causal knowledge obtained from process tracing focuses particularly on causal mechanisms (Runhardt, 2015; Waldner, 2012). Process tracing is easy to square with standard logics of causal knowledge under this concept. If one posits a mechanism under which some events in the process cause some other events in the process, one has explained the joint occurrence of the events. As in all theories, the many events one might have observed but that are not part of the causal process are implicitly held to be irrelevant. In addition, the mechanism embeds a counterfactualist logic, such that we can explain why the outcome in the case would have been different had some of the events been different. Whatever the theoretical tradition and behavioral postulates employed, the postulation of a causal mechanism to account for events in a single case *is* process tracing. Thus, HPE scholars aiming to explain the

---

[15]Even if technically possible, it is not generally desirable to include case facts in a causal mechanism that are not necessary to explain the outcome. First, irrelevant factors undermine the clarity of theoretical explanation within the case. Second, extraneous factors inhibit the application of the mechanism to other cases that may not match those factors.

reasons for a single sequence of events (or evaluate such an explanation) would be well advised to learn more about it.

Many scholars agree that while mechanisms are not observable, they do have observable implications or "markers." When aimed at causal mechanisms, a prototypical process trace inverts observable markers to set-identify a mechanism. That is, it "rules out" certain mechanisms as compelling explanations, and shows that any compelling explanation must lie in a particular set of mechanisms (possibly a singleton if the mechanism can be uniquely identified).

Many process tracing scholars also contend that it offers a special kind of causal leverage not found in other empirical methods (Collier *et al.*, 2004). This is sometimes summarized by asserting that process tracing studies something called a "causal process observation," as distinct from a "dataset observation" that one finds in datasets for statistical analysis. The causal inference we can draw from case evidence depends on whether the observable markers *may* or *must* occur if a given mechanism operates, whether certain markers can occur *only if* if a given mechanism operates, or both. Let μ denote a causal mechanism and m denote an observable marker. Then the possible inferences from the presence of a marker break down as follows.

1. m is *more likely* to occur if μ operates: μ raises the likelihood of m, but m is neither necessary nor sufficient for μ. In process tracing this is sometimes known as a "straw in the wind" test for μ (Collier, 2011; George and Bennett, 2005; Van Evera, 1997). For example, perhaps μ generates marker m, but the marker may not be recorded. So lack of occurrence of m could mean that μ does not operate, or that records of m were destroyed or never created. This is a particular problem in HPE when m relates to the beliefs or intentions of individuals. The statement is like, "If μ operates *and* records were kept, then m occurs." When m does not occur, it can mean μ does not operate, records were not kept, or both.

2. m *must* occur if μ: μ is a sufficient condition for m to occur; if μ operates, there is no way for m not to occur. In the HPE context this implies a strong statement about completeness of records. In this condition, observance of m confirms with certainty that μ is at least potentially operating, but it does not nail down μ because other mechanisms μ' might generate m as well. In process tracing this is sometimes known as a "hoop test" for the operation of μ.

3. m can occur *only if* μ: m is a sufficient condition for μ, so whenever m occurs we know μ operates. In this condition, m may not always occur even if μ operates, but if m occurs we infer with certainty that μ operates. This means that m cannot occur by any other mechanism

than μ. In process tracing this is sometimes called a "smoking gun" test for the operation of μ.

4. *m must* occur if μ, and *only if* μ. This is a combination of the previous two conditions, sometimes known in process tracing as a "doubly decisive" test for μ. Under this condition, the operative mechanism μ will be uniquely identified with certainty: it must generate m, and we will know on seeing m that μ generated it.

The key point about these conditions is that which one(s) are applicable depends on two things: a theory of how *all* mechanisms work, and the state of record keeping in the case. It has nothing to do with how the case evidence is collected, processed, or analyzed by the researcher. For a given marker m, it is not in the researcher's discretion to apply a smoking gun test or a doubly decisive test. If the set of conceivable mechanisms does not generate an "only if" statement for any marker then these tests cannot be used. Neither a research method nor a researcher determines whether a case is a "causal process observation." The set of theories potentially applicable to the case do.

More specifically, how does one know if a "smoking gun" logic is applicable? One must think not only about μ, but also about *every other conceivable mechanism*, and show that all mechanisms besides μ can never generate m.[16] This is a theoretical argument that cannot be demonstrated empirically because it relates to every possible theory one could apply, not the events that actually took place.

This has two implications about the smoking gun logic in process tracing. First, its logic of inference about mechanisms is strikingly similar to the logic of causal inference in instrumental variables regression. The "only if" statement is known in IV as an "exclusion restriction." In IV, an instrument is a variable Z that is unrelated to outcome Y except through Z's relationship to the factor X. This means that Z provides evidence about the causal effect of X on Y. The statement "unrelated except through" is the exclusion restriction. Substitute m for Z and μ for X, and we have a "smoking gun" relation between mechanisms and markers.

Second, in both process tracing and IV, demonstrating the "only if" condition is an extraordinarily tall order. In quantitative applications, my sense is that scholars have rightly become fairly critical of exclusion restriction arguments in all but the clearest cases of randomized assignment (e.g. experiments) or exogenous manipulation of assignment (e.g. sharp regression discontinuity). Except for those cases it is not clear how one goes about thinking

---

[16]This issue is not mitigated by the concept of equifinality, or multiple causal pathways (under one mechanism or distinct mechanisms) to a given outcome.

through the implications of every possible theory about the existence of a given marker m.[17]

To address this problem, process tracing usually involves "contrasting the observable implications of several alternative mechanisms" (Runhardt, 2015). This is important because the centerpiece of social science's contribution to historical knowledge is to understand the range of causal mechanisms that can account for a given case. Arguing that some particular mechanism cannot work through some particular events is informative.

However, for theory-from-data applications in social science, it would take a truly extraordinary theorist to claim to have evaluated every possible theory, because there are infinitely many of them. Every observable event in a process trace constitutes another needle to thread in a causal pathway. But note again that events and conjunctions in any process trace are inductively defined *by the process tracer*, not objectively given.[18] Given that, and given the infinite depth of historical experience (Weber, 1949), we can always find a new causal pathway to "the outcome" and construct a mechanism that works through it. Thus, rather than claiming causal uniqueness, we are better off claiming what is actually true: that case evidence is consistent with these mechanisms, not consistent with those mechanisms, and future work will have to figure out the rest.

Establishing the "Must" condition ("hoop tests") is only somewhat less demanding. It means that if μ is the causal mechanism, there is no possible realization of the case that lacks m. I suspect few scholars in HPE in particular would be willing to stake their claims on the assumption that there are *no* important stochastic elements in record keeping, yet that is what the "must" condition amounts to. If the existence of records noting the marker m is in part stochastic, it is not clear how we can justify the claim that m must be observed whenever μ operates.

The "straw in the wind" test is on more solid footing. It requires that, in the theory, there is a correlation between m and the operation of μ; this is what it means for m to be more likely when μ operates than when μ does not operate. This requires only that (i) μ does conduce to occurrence of m, and (ii) some other conceivable mechanism does not conduce to occurrence of m. The second condition is important: if every mechanism conduces to occurrence of m, then its occurrence is not informative about μ in particular.

---

[17]This is not to say it is impossible. Collier *et al.* (2004) provide an example of how cosmic microwave background radiation was discovered by accident from radio telescope interference. Astronomers realized that the big bang could explain CMBR, and nothing else reasonably could, so this became an important piece of evidence for the big bang. Importantly for my argument, it was theoretical knowledge of how the marker of radio interference could be produced under various mechanisms, not empirical evidence, that delivered this insight.

[18]Lustick (1996) argues that even the most discerning process tracers identify causal factors in view of their prior theoretical commitments, making it particularly problematic to claim that the constructed narrative corroborates the theory.

On the other hand, "straw in the wind" tests are the least compelling tests that process tracing offers for causal inference.

In light of all this, on what basis does process tracing provide unique causal leverage? The strongest leverage comes from "only if" conditions on observable markers of a mechanism. But as noted these conditions inhere in theory, not in empirical data of any kind or any method of analysis. It is essentially the same leverage one requires (or at least assumes) for instrumental variables regression. Thus for given evidence, causal leverage derives from neither "large n" nor "small n" structure, but "zero n" logic: it depends on theory about how various mechanisms work. This theory cannot be inferred from empirical evidence under any empirical testing strategy, because the theory is required to validate those tests.

The upshot of all this returns us to the primacy of theory in causal explanation. Because theories are necessary to structure the kinds of inferences it is valid to make from data, those inferences cannot be used to validate the said theories. This is particularly important for attempts to uniquely identify a mechanism from a case. Unique identification is impossible without an "only if" condition, and an "only if" condition is a strong statement about how various theoretical mechanisms work. On the other hand, "may" statements about observable markers are easier to sustain and provide information about the range of mechanisms consistent with evidence. If a mechanism implies the existence of a particular marker while some other mechanisms do not, then its occurrence is harder to square with the latter theories than the former. In short, in most cases, process tracing (like every approach to mechanism recovery) will work when it makes modest statements about a specific mechanism that is a candidate to explain a case, but not if it makes strong statements about a unique mechanism that must explain it.

Narratives, Analyticm and Otherwise. When historians build a causal explanation of an event, they often develop a narrative. Many political economists have little training in constructing or even reading historical narrative as such. The consequence may be that we are inclined to treat historical narratives as un-theoretical arrangements of facts lying inert to be explained, which is a mistake (Lustick, 1996). A narrative is a chronological presentation of events arranged in the form of a "story" to reach a conclusion (Danto, 1985). Thus like process tracing it involves the passage of time from the perspective of actors within a case. For some philosophers of history,[19] a narrative is inherently an argument of how the conclusion came about, and thus embeds a *causal* story (Carroll, 2001; White, 1967). In the usage of positive theory defined above, that qualifies as a theoretical statement. It is not news to narrative-builders that "every narrative is a selection of facts relevant to a specific theoretically informed investigation of a set of problems or events" (Parikh, 2000, p. 683).

---

[19]Not all: cf. Mink (1987); Winch (1958);.

Even for a narrative presented in a purely descriptive manner, the theoretical content lies in the selection of which events to include: these events are held to play a role in bringing about the conclusion.

Of course, most historical narrative by historians contains considerably deeper theoretical content. This content involves a selection from the "infinite depth" of historical experience (Weber, 1949). Traditional narrative draws out certain individuals as the key actors in a case; identifies, imputes, or implies intentional states such as goals and beliefs; and attempts to understand decisions in this way. Such a narrative does not state that other actors did not exist; it states that they are not necessary to understand how an event played out. That is the same kind of statement that social science theories make (Lustick, 1996): there are plenty of factors that a participant in the case might have noticed, but we do not need them to explain the outcome.

A key point for HPE is that historical narrative often builds an implicit model to explain a sequence of events producing a conclusion. This is the point of departure for "analytic narrative" (AN) (Bates *et al.*, 1998), an approach which preserves the narrative structure but makes the underlying model explicit. AN rests on identifying key facts in a historical case or cases, developing a model to fit them, and presenting a "narrative" of the case with the model as an underlying causal structure. The models are usually extensive form games, such that "early" decisions in a case generate a cascading causal pathway through "later" decisions.[20]

The theory embedded in a narrative is one account consistent with the case. There is no single narrative of any social event, and narrative historians have no general anxieties about causal uniqueness. However, scholars have criticized AN in particular as a social science method because it cannot uniquely identify the underlying model generating the case; it would seem that the same critique applies to standard narrative as well. As Clarke and Primo (2012) put it, "It may seem that there is little to be gained from using models to understand a

---

[20]There are several ambiguities about exactly what constitutes an AN, and more ambiguity about the stakes for whether a given study does or not. Some have referred to AN as "rational choice history" (Elster, 2000) while the progenitors note that they could be based on other behavioral postulates (Bates *et al.*, 2000), though this does not seem to have occurred in the literature to date. To some practitioners of AN, it "counts" to develop a model explaining a given case coupled with a narrative of that case (e.g. Defigueiredo Jr. *et al.*, 2006), but to others, a model-plus-case "does not an analytic narrative make" (Levi and Weingast, 2016, p. 2). In the latter conception, AN requires a model and a case, and also comparative statics derived from the model that are in principle testable. But unless actually testing the comparative statics is part of the definition of AN (in practice it is not defined or used with this restriction), essentially any model-plus-case counts because essentially all models have parameters and thus comparative statics that are in principle testable. A final requirement seems to be that ANs are generalizable: "The causal mechanisms and the structures or relationships must be generalizable to other cases under specifiable conditions" (Levi and Weingast, 2016, p. 6). Again, insofar as all models embed causal mechanisms and generalize to other cases within their equivalence class, this requirement seems to be met by any model-plus-case.

single event. After all, several different models may be written down that 'fit the data'" (p. 92).[21] A closely related criticism is that AN, and it would seem all theory-from-data accounts, are "Just So Stories" (Elster, 2000). Exactly what "just so story" means depends on the critic. It can mean an explanation that applies to only one case; an explanation designed specifically to fit certain data, and therefore not refutable by that data; an explanation that depends on a mechanism unlike any other mechanism that has ever operated; or an explanation based on a mechanism that is absurd. In any case, the "just so story" epithet is never a compliment.

In my view these criticisms miss the point. First, it is obvious that the case disciplines the explanation. One need only contemplate all the theories one might have constructed that do not fit the facts; we only read about the ones that do. It is true that an underlying model cannot typically be uniquely identified from a case, but it can be set-identified. This is still informative: from a positivist perspective, social science's empirical contribution is to understand the range of mechanisms consistent with available case evidence. Moreover, in a nontrivial case the array of evidence to be accounted for is substantial, which significantly disciplines the set of candidate models. Second, although multiple models can usually account for a given case, from an interpretive perspective it is useful to build them to show exactly how the case can be made intelligible. Third, like process tracing, one of the major points of narrative is not theory testing, but theory development. In a field where the mathematical truths of politics often seem to explain zero things well, it is a step in the right direction to explain one thing.

Still, if one wants the "best" explanation of a case, the issue is how to adjudicate among candidate explanations. The fact we must face is that any internally consistent account of events in a single case is a candidate to explain it. Stipulating two explanations that account for case evidence equally well, that evidence cannot adjudicate among them: they are observationally equivalent with respect to that case.

There are three ways to address observational equivalencies. First is to apply "strong inference" (Platt, 1964) on additional within-case data: identify points where the theories are not observationally equivalent, identify new data on those points, and evaluate each candidate's ability to explain it. The implicit point is that the JSS critique assumes an impoverished concept of a "case." In practice, a case never consists of a fixed body of known facts. There are always more information in other secondary sources, extant primary sources, and primary sources not yet discovered; historical experience contains infinite depth. Any explanation of a case points to key determinants, which can spur the search for additional data on them. In short, when a model is

---

[21]I will argue below that this also applies to every possible interaction between models and data in social science (cf. Demarchi, 2005, p. 14), so it is not clear why it matters particularly for theory-from-data approaches.

derived from a case, it becomes "testable" against additional evidence that can be adduced from the case, which is essentially always possible.

The second response to observational equivalencies is to consider theoretical generalizability. How easy is it to imagine applying the same theoretical mechanism in a different case or substantive context, or how specialized does it seem for the specific case at hand? If a mechanism depends on a long conjunction of causal factors, some of which have never been observed outside the specific case, it is unlikely to make an important contribution to the library of mechanisms in political economy.[22]

The third response is to evaluate how well the candidate theories make the decisions and outcomes in the case intelligible. To evaluate intelligibility it is useful to evaluate model assumptions. Are the goals, constraints, tradeoffs, information asymmetries, rivalries, exogenous shocks, etc. imputed by the model at all veridical?[23] Evaluating assumptions is a long debated proposition in science, with recent defenses of it for certain purposes in formal political theory (Lorentzen *et al.*, 2017; Paine and Tyson, 2020). In my view its value is not about science but about interpretation.

Overall, if all historical narrative embeds a theory, then the only question is whether the theory will be explicit or implicit, highlighted for engagement and critique by the reader or left between the lines. While much historical narrative works perfectly well at conveying meaning about the causation of events, there is a place for more explicit and formal statement of the theoretical content underlying a narrative argument.[24]

Theory-from-Data: Summing up. In theory-from-data analysis, we identify a mechanism consistent with some event, and in that sense a candidate to explain it. Consider again Lizzeri and Persico (2004). From this paper we learn a precise depiction of a causal mechanism that is consistent with voluntary franchise extension in 19th century England (and based on historical narratives of that development), and numerous other aspects of policy and party development in this case. This mechanism is also clearly a new entrant in the political economy library. Is this the unique causal mechanism to explain the case? Clearly not; Acemoglu and Robinson (2000) present just one competing explanation. Competing explanations do not explain the conjunction of events cited by Lizzeri and Persico; they implicitly hold that these events are ancillary.

---

[22]This theoretical generalizability is different from the empirical generalization highlighted by many (e.g. Bates *et al.*, 2000). Empirical generalization means that a causal mechanism actually does induce effects in a wide variety of cases. I discuss this further below.

[23]This is importantly different from claiming that every perception by a participant in the case reflects relevant causal information or should be reflected in a model.

[24]As Carpenter (2000) notes, historical narrative contains elements that formal models cannot easily replicate. We do not need to try in order for explicit theoretical models to add something useful.

Does the causal mechanism generalize to other cases? We do not learn that either.

Nevertheless, the model provides a basis for comparison with other cases. Theory-from-data work will always face the question of whether our interest is "in finding some general lawlike statements or in explaining a particular event" (Beck, 2006, p. 349). Theories-from-data do both. Being tailored to a specific event, they are constructed as an explanation of it. At the same time, models inherently capture a generalizable logic reflected in their equivalence classes and mechanisms. They define a general class of situations, and apply to any case in that class. It may so happen that the class has only one known member in historical data, but the possibility remains that further research will turn up others.

Taking Theory to (Historical) Data

Evaluation of theory based on data follows a variety of specific paths but most share a similar (and familiar) structure. It starts with a set of hypotheses implied by a theory. Hypotheses might assert that a set of factors X and the outcome Y always jointly occur, or that certain factors do not occur in the same case as the outcome Y. Such hypotheses apply when certain factors are necessary or sufficient for the outcome, and conduce to cross-case qualitative analysis. Hypotheses might assert correlational statements such as more of X is associated with less of Y. Such hypotheses correspond naturally to comparative statics from models and are the norm in regression modeling.

Whatever the mode of analysis, it informs us of the consistency between the hypotheses and the data. For instance, a p-value in a statistical hypothesis test answers the question, "On a scale of 0–100 (100 being the easiest), how easy is it to explain the observed data if the hypothesis is wrong in a specific way?" Qualitative methods yield less formalized output, but (no pun) qualitatively similar findings about the correspondence of theoretical expectations and empirical observations. Thus, despite real differences highlighted in the "two cultures" discourse in political methodology (Goertz and Mahoney, 2012), there is a broad similarity in both the questions asked and information conveyed in both quantitative and qualitative investigations (Lorentzen *et al.*, 2017).

What do we learn from this? We learn something about the consistency of the theory and the case(s) from which the data was drawn. If the hypotheses are consistent with the data, the theory is a candidate to explain some part of the case.

However, as with theory-from-data inference, causal uniqueness is a different story. We cannot gain *empirical* knowledge that the theory is the only candidate to explain the data, i.e. data cannot imply that a specific causal mechanism

is actually working in the case that generated the data.[25] Such a conclusion requires that the mechanism in question is the only one that can generate the hypotheses. As I noted above about process tracing, we are very rarely in a position to make that claim.[26] Even if we can, causal uniqueness is a *theoretical* claim about the operation of all possible mechanisms.[27]

Dippel *et al.* (2020) present an exemplar of the theory-to-data approach. They develop a model of employers lobbying a government to enact coercive policies that reduce outside options for laborers. In an agricultural setting the key comparative statics are that greater planter power should be accompanied by more coercive policy and lower wages. They investigate these relationships in panel data from the British Caribbean, where planter power is operationalized as the share of plantation output in total exports. An instrument for this factor is used to argue that planter power *causes* higher incarceration rates and thereby lower wages, which is clearly consistent with the causal mechanism in the model.[28] Is it consistent with others? Usually when we search for other mechanisms we argue that the identification strategy has failed in some way, which re-opens the door for confounding or reverse causation. But suppose we stipulate that the IV and mediation analysis are credible. Then there are still other causal explanations to account for the evidence. For just one example, consider the social destruction wrought by a plantation economy, the effect of this on antisocial behavior, and thus incarceration, human capital development, and wages. The evidence in the paper is also consistent with this alternative mechanism. This point is not a critique of the paper, which is admirably careful about investigating its causal mechanism, given data limitations, and successful in casting doubt on a wide range of competing explanations. The point is that even with all this care, we still can only conclude that one candidate mechanism is consistent with empirical evidence in this setting.

In short, in both theory-from-data and theory-to-data approaches, the essential contribution is to understand the consistency of a causal mechanism

---

[25]Clarke and Primo (2012) strongly criticize the standard practice reviewed here, but it seems that sticking to claims that a specific theory is a candidate to explain the case, rather than confirmed as the unique explanation, satisfies most of their objection.

[26]This is particularly important for theories asserting X causes Y, and that imply joint occurrence or correlation between certain variables. Those observations are also consistent with reverse causation (Y causes X) and confounding (some unmeasured Z causes both X and Y).

[27]In addition, internal consistency of theory and evidence in any set of cases does not tell us anything about the theory as such. It does not tell us whether the theory is "true" or "false." A conclusion of this nature from empirical findings does not make sense. The theory in itself is a series of logically related postulates and conclusions not subject to empirical verification or disconfirmation (cf., e.g., Clarke and Primo, 2012).

[28]A separate mediation analysis argues that, assuming the causal model is correct, most of the effecet of plantation export share on wages is mediated by coercion, which is also consistent with the model.

with a given collection of case evidence, and possibly to add new mechanisms to our library. For these contributions it makes no difference whether the theory or the data comes first. But in neither approach is it possible to make an empirical claim that we have found the unique or best causal explanation. This claim is theoretical: it requires considering every possible mechanisms consistent with the evidence.

Theory and Credibility in HPE

For many empirical scholars in HPE, the operational goal in research will be credible identification of a treatment effect.[29] Credible identification cuts across the theory-from-data and theory-to-data distinction. In the theory-from-data vein, a "found" randomization or discontinuity can be used to identify a causal effect, and then a theory developed to explain this finding. Credible identification obviously also works with theory-to-data applications. Starting with a theory that X causes Y, one wants a case where X is randomly or discontinuously assigned, etc. and Y is observed.[30]

In practice, when we see a paper with a theory and empirics in a "found" randomization, it is often difficult for anyone to know whether the theory or the randomization came first (e.g. consider Guardado (2018) on colonial corruption and long-term development in Peru; Cirone and Van Coppenolle (2019) on lottery procedures and institutional development in the French Third Republic; and Garfias and Sellars (2021) on demographic collapse and state centralization in colonial Mexico). As I argued above, it does not matter. In either case, we learn that mechanisms in a certain subset of our library are potentially good candidates to explain the observed data, and mechanisms in another subset are not good candidates.

However, in either direction between theory and data, credible research designs are particularly valuable in causal explanation because they restrict the set of reasonable mechanisms more than most research designs. They do this by creating a tighter connection between the all-else-equal requirement of causal theories and the empirical findings (Ashworth *et al.*, 2021). Thus "positive" findings of a treatment effect tell us that mechanisms of reverse causation and confounding are not good candidates to explain some given data; "negative" findings tell us that a mechanism of causation is not an especially compelling explanation. By contrast, with a correlation that does not credibly identify any treatment effect, mechanisms of causation, reverse causation, and

---

[29]By "credibility" I mean a credible exclusion restriction in empirical analysis, such as randomization or discontinuities in assigning factors to cases, which identify a treatment effect of those factors in that case: the amount of change in Y caused by a 1-unit change in X.

[30]This usually requires selection of cases to meet the demand of the method; I will argue below that in HPE this creates more problems than it solves.

confounding are all potentially reasonable candidates to explain the observed data.

Thus, more credible designs are thus more informative about the operative mechanism than less credible designs. However, identifying a causal effect is different from identifying a causal mechanism. Any mechanism that explains the treatment effect is a candidate explanation. Credibility notwithstanding, there will always be multiple candidates, for the same reasons discussed above. Given multiple mechanisms that explain the treatment effect, evaluating them is not an empirical matter, but a theoretical one.

Causal Generalization and Historical Knowledge

Social scientists usually do not only want to offer a candidate explanation for specific cases. We want *causal uniqueness* within a case, and *causal generalization* to a broad range of cases. I have already made my argument against any *empirical* demonstration of causal uniqueness so I focus here on causal generalization. I will argue that any empirical causal generalization is necessarily inductive, and any non-inductive causal generalization is both purely theoretical and not typically convincing.

The issue in causal generalization is whether or when we can transfer knowledge of a causal mechanism empirically identified in one case to another case *without executing the same empirical analysis of the second case.*[31] The caveat is important. We can of course replicate empirical analysis across a variety of cases and examine how well a causal mechanism accounts for a variety of cases. This produces inductive knowledge of causal mechanisms. The causal generalization question that I address is how to know that a causal finding in a case transfers to others, simply by virtue of its identification in the original case.[32]

For instance, suppose we find a relationship between economic structures implemented by colonizers and long-run productivity in a particular case (Dell and Olken, 2020). Inasmuch as this finding pertains to historical legacies of an important case, and the present living conditions of a large number of people, it need not generalize to any other case to be important. Nevertheless, social scientists often wish to apply such findings to other cases. On what basis could do so? We do so when we see a similarity to the original cases in a way that is relevant under some theory, so we suspect that a similar mechanism operates. If the mechanism delivers a specific relationship between colonial economic

---

[31]Thus causal generalization is related to *external validity.* One cannot properly speak of causal generalization until one has solid local causal knowledge, just as external validity cannot exist without internal validity. But as I will argue, the fundamentally theoretical nature of external validity claims holds even when the highest standard for internal validity is met.

[32]I believe that this argument is the linchpin of what one learns from well-identified studies in cases that are not themselves of any particularly broad interest.

structures and long-run productivity in the original study, we could reason that it implies a similar relationship in similar cases.

Somewhat more formally, consider the causal generalization argument (CG): if (i) a mechanism μ is credibly identified in case A, and if (ii) cases A and B are both members of μ's equivalence class, then (iii) μ can be inferred to operate in case B. Thus, one can make credible causal statements in case B without a direct empirical study of that case, on the basis of credible identification in case A.

Two cases, A and B, in μ's equivalence class get there by having similar factors X_A and X_B that have causal capacities under μ. Call this *factor similarity*. The magic of the CG argument is to translate factor similarity, which is more or less observable, into what we may call *mechanistic similarity*. Mechanistic similarity means that the same causal mechanism operates in two cases.

The CG argument is the keystone of any non-inductive generalizability of empirical findings. The key question is how to evaluate step (ii). Without actually executing an empirical study in case B, step (ii) works only if we assume that factor similarity across cases will lead to similar outcomes across cases. This means that we know how to translate a relationship between X_A and Y_A into a relationship between X_B and Y_B, despite not actually observing this relationship. That is, we can understand the unobserved relationship between X_B and Y_B — identify a candidate causal mechanism in case B — simply because we know the relationship between X_A and X_B (Pearl and Bareinboim, 2014).

This translation is a special case of a research design we know by another name, selection on observables. Selection on observables means that, if we know an outcome Y and treatment status X in one case, we can infer a counterfactual outcome Y' under an alternative treatment X'. This is possible if and only if we know the structural model that translates X into Y, which per se entails the causal effect of X. Moving from X_A to X_B across two cases presents the exact same problem, and solving it requires the same assumption.[33]

We get a lot of mileage in generalized causal explanation from selection on observables, provided we believe it. The next question is, if selection on observables is valid at step (ii) of the CG argument, why is not it valid at step (i)? If it is, then no specialized research design such as experiments, RD, and the like is necessary to establish credible causal statements at all. If it is not — and as credibility revolutionaries most of us probably agree it usually is not — then credible causal identification in case A tells us nothing about

---

[33]My argument is not that the CG argument is vacuous. It holds for example when cases are randomly sampled in a purposive, controlled sampling design from a larger universe of cases. In this instance, causal findings from a sample obviously generalize to the population of un-analyzed cases. In HPE, we typically will not sample cases or control a sampling process in this way.

case B that we would not already know from a theory that μ operates in case B. In the real-world conditions that I suspect are almost always relevant for HPE, empirical causal generalization is self-abnegating: it assumes the exact knowledge that we are seeking to establish.

Put somewhat differently, causal mechanism generalization from case A to case B requires a compelling mechanism to account for an effect in case A, and a theory of case B that says it is similar to case A in all respects relevant to the mechanism. But if we believe the theory required for the second step, we already know that the mechanism operates in case B. Causal identification in case A adds nothing to understanding of causal mechanisms in case B. On the other hand, if we do not believe the theory required for the second step, then we do not know how to translate the causal effect in A to an effect in B. Causal identification in A again adds nothing to understanding of B.

It is at this point that many scholars discover their inner Bayesian. After all, if we learn that a mechanism operates somewhere, surely we should increase our confidence that it operates somewhere else, all the more so if the cases are similar in obvious ways (Ashworth *et al.*, 2021). This argument is fine as far as it goes, but it sits uncomfortably with our demands for within-case analysis. One cannot be Bayesian for causal generalization but an identificationist for within-case causal explanation, because they are the same problem.[34] If we accept the credibility revolution's standards for causal statements, the best we can hope for in causal generalization is that explaining a case leads to a qualitatively new causal mechanism in our library, which might help explain other cases.

In light of all this, consider again the analysis of Dell and Olken (2020) on the positive effect of colonial economic structures on long-run productivity. The causal effects are carefully identified from Dutch colonization of Java in the 19th century. The paper notes two key channels plausibly responsible for the effect — development of manufacturing and colonial infrastructure investment (roads and rail) — that could be developed into a theory. All of its arguments are appropriately localized to this case. Beyond the intrinsic import of Javanese development, we may ask: what do we learn *in general* from this analysis? Empirically, nothing. It provides no basis to believe (and does not claim) that other places where colonizers developed manufacturing and infrastructure have also experienced long-run productivity gains, because we have no way to know, in the absence of an assumption or a separate study, whether those other cases are similar enough to Dutch colonization in Java or postcolonial governance there. In highlighting candidate mechanisms, the paper suggests that factors scholars might consider in other cases. From this

---

[34]Moreover, the allowance for Bayesian updating from biased observational research undermines the hard core identificationist position (Little and Pepinsky, 2021).

we might inductively develop general causal understanding. But the findings of this paper cannot extend beyond its case.

Causal Localism and Case Selection

To summarize, theory-from-data and theory-to-data approaches make qualitatively the same contributions to historical understanding. With respect to any case or set of data, they tell us whether a specific mechanism is a candidate to explain it. For general understanding, the contributions are new entries in the "library of mechanisms" that scholars can deploy to understand other cases. However, we cannot say that finding a causal mechanism operates (or has a particular effect) in one case tells us that the mechanism operates in another case unless we already possess *a priori* theoretical knowledge of the causal structure of the second case. But armed with that knowledge, the empirical finding from the first case is irrelevant; we would already know what we need to know about the second.

If all empirical causal knowledge is local, then the only question in selecting cases for empirical study is whether we want to know something about them. We should want to know something about cases either when they are historically important, or when they seem to require explanation with new causal mechanisms not already in our library. If we study these cases, we will learn something about history and we might learn something about theory (a new entrant in our library). But it is no use studying a case simply because it leads to clean identification of something. Whatever mechanisms operate there cannot by virtue of that identification be said to operate elsewhere. Whatever is learned empirically is local to that case. It is also no use declining to study an interesting case simply because various identification problems are challenging. As we constitute this field, we should study important cases and flesh out the (possibly large) range of mechanisms that can explain them, not select our cases on the basis of what can be understood in a particular way.

I do not intend this argument to denigrate the generalized causal understanding obtainable in HPE (or for that matter in empirical social science, because the same argument applies to most of it). A large library of mechanisms is useful. It gives us a powerful toolkit for identifying factors to watch for in new cases (provided we guard appropriately against confirmation bias). It gives us a supple language for the grouping of like cases and the distinguishing of unalike ones following empirical analysis of those cases. This is valuable, and moreover it is the only generalized causal knowledge we can expect to develop in HPE. So we should embrace it.

Conclusion

Historical political economy depends fundamentally on theory for both its constitution as a field and for the success of its empirical work. Theory is necessary for causal explanation, and explanation is necessary to understand the significance of empirical findings, both in historical and social scientific terms. In addition, I have argued that essentially all statements we will be able to muster in this field about either the uniqueness of causal explanation within a case, or generalization of causal explanation across cases, are entirely theoretical and not empirical in nature. Empirical investigation that either "tests" or "derives" theory (i.e. theory-from-data and theory-to-data) informs us about the consistency of causal mechanisms with a specific body of evidence. Regardless of the approach, and whether one follows its qualitative or quantitative idiom, empirical research presents new entries in a "library of mechanisms" that comprise HPE's contribution to generalized historical knowledge.

References

Acemoglu, D., S. Johnson, and J. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation". *The American Economic Review* 91(5): 1369–401.

Acemoglu, D. and J. Robinson. 2000. "Why Did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective". *Quarterly Journal of Economics* 115(4): 1167–99.

Acharya, A., M. Blackwell, and M. Sen. 2018. *Deep Roots: How Slavery Still Shapes Southern Politics*. Princeton, NJ: Princeton University Press.

Alesina, A., B. Reich, and A. Riboni. 2020. "Nations, Nationalism, and Wars". *Journal of Economic Growth* 25(4): 381–430.

Ashworth, S., C. Berry, and E. Bueno de Mesquita. 2021. *Theory and Credibility*. Princeton, NJ: Princeton University Press.

Bates, R. H., A. Greif, M. Levi, J.-L. Rosenthal, and B. R. Weingast. 1998. *Analytic Narratives*. Princeton University Press.

Bates, R. H., A. Greif, M. Levi, J.-L. Rosenthal, and B. R. Weingast. 2000. "The Analytic Narrative Project". *American Political Science Review* 94(3): 696–702.

Beck, N. 2006. "Is Causal-Process Observation an Oxymoron?" *Political Analysis* 14(3): 347–52.

Black, M. 1962. *Models and Metaphors: Studies in Language and Philosophy*. Cornell University Press.

Blatter, J. and M. Haverland. 2012. *Designing Case Studies: Explanatory Approaches in Small-N Research*. Palgrave Macmillan.

Cantoni, D. and N. Yuchtman. 2020. "Historical Natural Experiments: Bridging Economics and Economic History". *Tech. rep.* National Bureau of Economic Research.

Carpenter, D. 2000. "Commentary: What Is the Marginal Value of Analytic Narratives?" *Social Science History* 24(4): 653–67.

Carroll, N. 2001. *Beyond Aesthetics: Philosophical Essays.* Cambridge University Press.

Cartwright, N. 1999. *The Dappled World: A Study of the Boundaries of Science.* Cambridge University Press.

Cirone, A. and B. Van Coppenolle. 2019. "Bridging the Gap: Lottery-Based Procedures in Early Democratization". *World Politics* 71(2): 197–235.

Clarke, K. A. and D. M. Primo. 2012. *A Model Discipline: Political Science and the Logic of Representations.* Oxford University Press.

Collier, D. 2011. "Understanding Process Tracing". *PS: Political Science & Politics* 44(4): 823–30.

Collier, D., H. Brady, and J. Seawright. 2004. "Sources of Leverage in Causal Inferece: Toward an Alternative View of Methodology". In: *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* Ed. H. Brady and D. Collier. Cambridge University Press.

Collingwood, R. G. 1946. *The Idea of History.* Oxford University Press.

Coşgel, M., T. Miceli, and J. Rubin. 2012. "The Political Economy of Mass printing: Legitimacy and Technological Change in the Ottoman Empire". *Journal of Comparative Economics* 40(3): 357–71.

Cox, G. 2004. "Lies, Damned Lies, and Rational Choice Analyses". In: *Problems and Methods in the Study of Politics.* Ed. I. Shapiro, R. Smith, and T. Masoud. Cambridge University Press, 167–85.

Danto, A. 1985. *Narration and Knowledge.* Columbia University Press.

Defigueiredo Jr., R. J., J. Rakove, and B. R. Weingast. 2006. "Rationality, Inaccurate Mental Models, and Self-confirming Equilibrium: A New Understanding of the American Revolution". *Journal of Theoretical Politics* 18(4): 384–415.

Dell, M. and B. Olken. 2020. "The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java". *Review of Economic Studies* 87(1): 164–203.

Demarchi, S. 2005. *Computational and Mathematical Modeling in the Social Sciences.* Cambridge University Press.

Diamond, J. and J. A. Robinson. 2010. *Natural Experiments of History.* Harvard University Press.

Dippel, C., A. Greif, and D. Trefler. 2020. "Outside Options, Coercion, and Wages: Removing the Sugar Coating". *Economic Journal* 130(630): 1678–714.

Dowding, K. 2015. *The Philosophy and Methods of Political Science.* Macmillan International Higher Education.

Dray, W. 1957. *Laws and Explanation in History*. Oxford University Press.

Dupré, J. and N. Cartwright. 1988. "Probability and Causality: Why Hume and Indeterminism Don't Mix". *Nous* 22(4): 521–36.

Elster, J. *et al.* 1998. *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge University Press.

Elster, J. 2000. "Rational Choice History: A Case of Excessive Ambition". *American Political Science Review* 94(3): 685–95.

Engerman, S. and K. Sokoloff. 2002. "Factor Endowments, Inequality, and Paths of Development among New World Economies". *Economía* 3(1): 41–109.

Falleti, T. G. and J. F. Lynch. 2009. "Context and Causal Mechanisms in Political Analysis". *Comparative Political Studies* 42(9): 1143–66.

Gailmard, S. 2019. "Imperial Politics, English Law, and the Strategic Foundations of Constitutional Review in America". *American Political Science Review* 113(3): 778–95.

Gailmard, S. 2020. "Game Theory and the Study of American Political Development". *Public Choice* 185: 1–23.

Gailmard, S. 2021. "Imperial Governance and the Growth of Legislative Power in America".

García Jimeno, C. 2016. "The Political Economy of Moral Conflict: An Empirical Study of Learning and Law Enforcement under Prohibition". *Econometrica* 84(2): 511–70.

Garfias, F. and E. Sellars. 2021. "From Conquest to Centralization: Domestic Conflict and the Transition to Direct Rule". *Journal of Politics* forthcoming.

Gennaioli, N. and H.-J. Voth. 2015. "State Capacity and Military Conflict". *Review of Economic Studies* 82(4): 1409–48.

George, A. L. and A. Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. MIT Press.

Gerring, J. 2006. *Case Study Research: Principles and Practices*. Cambridge University Press.

Goemans, H. and W. Spaniel. 2016. "Multimethod Research: A Case for Formal Theory". *Security Studies* 25(1): 25–33.

Goertz, G. 2017. *Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach*. Princeton University Press.

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

Gordon, S. and H. K. Simpson. 2020. "Causes, Theories, and the Past in Political Science". *Public Choice* 185: 315–33.

Guala, F. 2005. *The Methodology of Experimental Economics*. Cambridge University Press.

Guardado, J. 2018. "Office-Selling, Corruption, and Long-Term Development in Peru". *American Political Science Review* 112(4): 971–95.

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

Hall, P. A. 2003. "Aligning Ontology and Methodology in Comparative Research". *Comparative Historical Analysis in the Social Sciences*.

Hedström, P. and R. Swedberg. 1998. "Social Mechanisms: An Introductory Essay". *Social Mechanisms: An Analytical Approach to Social Theory*: 1–31.

Imai, K., L. Keele, D. Tingley, and T. Yamamoto. 2011. "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies". *American Political Science Review*: 765–89.

King, G., R. O. Keohane, and S. Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton University Press.

Levi, M. and B. Weingast. 2016. "Analytic Narratives, Case Studies, and Development". *Tech. rep.* SSRN.

Little, A. and T. Pepinsky. 2021. "Learning from Biased Research Designs". *Journal of Politics* forthcoming.

Little, D. 1991. *Varieties of social explanation*. Vol. 141. Boulder: Westview Press.

Lizzeri, A. and N. Persico. 2004. "Why did the Elites Extend the Suffrage? Democracy and the Scope of Government, with an Application to Britain's 'Age of Reform'". *Quarterly Journal of Economics* 119(2): 707–65.

Lorentzen, P., M. T. Fravel, and J. Paine. 2017. "Qualitative Investigation of Theoretical Models: The Value of Process Tracing". *Journal of Theoretical Politics* 29(3): 467–91.

Lustick, I. 1996. "History, Historiography, and Political Science: Multiple Historical Records and the Problem of Selection Bias". *APSR* 90(3): 605–18.

Machamer, P., L. Darden, and C. F. Craver. 2000. "Thinking about Mechanisms". *Philosophy of Science* 67(1): 1–25.

Mandlebaum, M. 1938. *The Problem of Historical Knowledge: An Answer to Relativism*. New York: Liveright.

McAdam, D., S. Tarrow, and C. Tilly. 2008. "Methods for Measuring Mechanisms of Contention". *Qualitative Sociology* 31(4): 307.

Mink, L. O. 1987. "Philosophical Analysis and Historical Understanding". In: *Historical Understanding*. Ed. B. Fay, E. Golob, and R. Vann. Cornell University Press, 118–46.

Myerson, R. 2006. "Game-Theoretic Consistency and International Relations". *Journal of Theoretical Politics* 18(4): 416–33.

Paine, J. and S. A. Tyson. 2020. "Uses and Abuses of Formal Models in Political Science". In: *The SAGE Handbook of Political Science*.

Parikh, S. 2000. "Commentary: The Strategic Value of Analytic Narratives". *Social Science History* 24(4): 677–84.

Pearl, J. and E. Bareinboim. 2014. "External Validity: From Do-Calculus to Transportability across Populations". *Statistical Science* 29(4): 579–95.

Platt, J. R. 1964. "Strong Inference". *Science* 146(3642): 347–53.

Runhardt, R. W. 2015. "Evidence for Causal Mechanisms in Social Science: Recommendations from Woodward's Manipulability Theory of Causation". *Philosophy of Science* 82(5): 1296–307.

Salmon, W. C. 1984. *Scientific Explanation and the Causal Structure of the World.* Princeton University Press.

Salmon, W. C. 1998. *Causality and Explanation.* Oxford University Press.

Satz, D. and J. Ferejohn. 1994. "Rational Choice and Social Theory". *The Journal of philosophy* 91(2): 71–87.

Sekhon, J. S. and R. Titiunik. 2012. "When Natural Experiments Are Neither Natural Nor Experiments". *American Political Science Review*: 35–57.

Shepsle, K. 2009. "Why?" In: *The Future of Political Science: 100 Perspectives.* Ed. G. King, K. Schlozman, and N. Nie. Routledge, 244–6.

Van Evera, S. 1997. *Guide to Methods for Students of Political Science.* Cornell University Press.

Waldner, D. 2012. "Process Tracing and Causal Mechanisms". *The Oxford Handbook of Philosophy of Social Science*: 65–84.

Weber, M. 1949. *The Methodology of the Social Sciences.* Free Press.

White, M. 1967. "Foundations of Historical Knowledge". *British Journal for the Philosophy of Science* 18(1): 72–4.

Winch, P. 1958. *The Idea of a Social Science and Its Relation to Philosophy.* Routledge & Kegan Paul.

Woodward, J. 2002. "What is a Mechanism? A Counterfactual Account". *Philosophy of Science* 69(S3): S366–S377.