Research Design in Political Science
POLS4011/POLS8058

Richard Frank
21 April 2026

WEEK 7: CAUSAL INFERENCE, PART 1 (THE PROBLEM)

PART 1: OVERVIEW

The previous six weeks of this class built a scaffolding: how to ask a research question (Weeks 1–2), how to define and structure the concepts in that question (Weeks 3–4), and how to select and scope the cases that will test it (Weeks 5–6). All of that was building toward the question that motivates most empirical political science: *can we establish that X caused Y?* Weeks 7 and 8 take that question head-on. This week establishes the problem; next week surveys the solutions.

The central message this week is that causal inference is hard, and it is hard for specific reasons not just because the world is complicated. The fundamental problem of causal inference is that we can never observe the same unit (person, country, etc.) in both the treated and untreated state at the same time. Every causal claim therefore requires us to reason about something we cannot see, the counterfactual. Once you can internalise this, the correlation-is-not-causation mantra stops being a vague warning and becomes a clear statement about what is missing from a simple comparison. And the various research design strategies we will discuss next week (e.g., experiments, natural experiments, difference-in-differences, instrumental variables, regression discontinuity) are all, in the end, different ways of constructing a credible counterfactual.

This week also introduces two different traditions of thinking about causation in political science. First, the potential outcomes framework, which is drawn from statistics and economics, dominates quantitative research and focuses on average treatment effects across populations. Second, the case-oriented tradition, drawn from comparative-historical analysis and philosophy, focuses on necessity and sufficiency in individual cases. These are not rival epistemological camps. Mahoney (2008) argues that they are complementary, but they do emphasise different things. Your own research design will likely draw more heavily on one approach than the other, but understanding the logic of the other will make you a better consumer of research that uses it as well as understand the limitations of your own approach.

Plan for today

- Overview: connecting Weeks 5 & 6 to Weeks 7 &8
- Readings: the problem of causal inference and causation in case-oriented research
- Group activity: diagnosing causal claims in published research
- Wrapping up and looking ahead to Week 8

Key themes for this week

- The counterfactual definition of causation: a cause is something whose absence would have led to a different outcome.

- The fundamental problem of causal inference: we cannot observe the same unit in both treated and untreated states, so individual causal effects are unobservable.
- Case-oriented versus population-oriented approaches to causation: necessity, sufficiency, INUS causes, and how they relate to mean causal effects.
- Confounding, endogeneity, and selection bias as obstacles to credible causal claims.
- The applied example: how Cho, Dreher, and Neumayer (2013) illustrate the difficulty of making causal claims from cross-sectional observational data.

The differentiated expectations continue. Honours students should be able to ***explain*** the fundamental problem of causal inference and ***identify*** why correlation does not imply causation in specific examples. MA/PhD students should be able to ***distinguish*** between case-oriented and population-oriented conceptions of causation and ***evaluate*** published causal claims in terms of the assumptions required to sustain them.

PART 2: READINGS

Mahoney, James. 2008. "Toward a Unified Theory of Causality." *Comparative Political Studies* 41(4/5): 412-436.

Morgan, Stephen L, and Christopher Winship. 2015. *Counterfactuals and Causal Inference*, 2nd ed., Cambridge University Press. Ch. 1 & Ch 2 (sections 2.1 & 2.2 only).

Cho, Seo-Young, Axel Dreher, and Eric Neumayer. 2013. "Does Legalized Prostitution Increase Human Trafficking?" *World Development* 41: 67–82.

Mahoney (2008)

This article is the main theoretical reading this week. Mahoney's (2008) goal is ambitious. He wants to show that the case-oriented and population-oriented traditions in political science are not, as they are often treated, fundamentally different ways of thinking about causation, but are instead connected by a shared logical structure. The article is dense and can be intimidating, but the core ideas are accessible and worth working through carefully because they provide a vocabulary for talking about causation that bridges the qualitative-quantitative divide.

Mahoney (2008) starts by observing that political scientists working in different traditions tend to define causation differently. Population-oriented researchers (quant folk) usually define a causal effect as the difference in average outcomes between treated and untreated groups. Case-oriented researchers tend to think about causation in terms of necessary and sufficient conditions. Did a particular factor have to be present for the outcome to occur, and was it enough on its own to produce it? These seem like very different questions, and indeed much of the methods debate in political science proceeds as if they are. Mahoney (2008) argues they are not.

The key theoretical tools Mahoney (2008) borrows are from J.L. Mackie. An **INUS cause** is an *Insufficient but Necessary part of an Unnecessary but Sufficient condition*. That is a mouthful! Put simply, an explanatory factor is an INUS cause of an outcome if it is one component in a group of conditions that together are sufficient to produce the outcome, and that factor is necessary within that group (without it, the group would not be sufficient), but the group itself is not the only way the outcome could occur. Democracy promotion by external actors, for instance, might be an INUS cause of democratisation. It is not sufficient on its own

(other conditions must also hold), and it is not strictly necessary (countries can democratise without it), but when it combines with other factors (e.g., domestic elites, economic crisis, a weakened military) the group may be sufficient.

A **SUIN cause** is the mirror image of INUS: *Sufficient but Unnecessary part of an Insufficient but Necessary condition*. A factor is a SUIN cause if it is one of several factors that are each individually sufficient to contribute to a necessary condition. This is less intuitive, but the key idea is that SUIN causes represent equifinality at the level of necessary conditions. There are multiple paths to satisfying a condition that must be met.

The payoff of Mahoney's (2008) framework is in showing how case-level and population-level claims relate. A mean causal effect (i.e., a regression coefficient) is, Mahoney (2008) argues, a "symptom" of an INUS cause operating in the population. When we find that a variable X has a positive average effect on Y across many cases, what this typically means is that X is a component of one or more sufficient conditions for Y, and those conditions are operative in enough cases to produce a detectable average association. The mean causal effect does not tell us that X caused Y in every case; it tells us that the causal group containing X was active often enough to shift the average. This means that when a quantitative researcher reports an average treatment effect, the case-oriented researcher can ask in which cases was X actually part of a sufficient condition for Y, and in which cases was the outcome produced by a different causal pathway (equifinality)?

Mahoney (2008) also introduces probabilistic versions of necessity and sufficiency. A strictly necessary condition must be present every time the outcome occurs; a probabilistically necessary condition is present in most but not all instances. Similarly, a probabilistically sufficient condition usually produces the outcome but not always. This is important because strict necessity and sufficiency are extremely demanding standards that few empirical claims in political science can meet. The probabilistic versions are more realistic, and they correspond naturally to the kinds of claims we make with statistical models: a variable that is statistically significant is, roughly speaking, a probabilistically necessary or sufficient INUS cause.

Whether you are doing a quant or qual project, Mahoney (2008) gives you a way to think about what kind of causal claim you are making. If you are running a regression, you are estimating mean causal effects. What does that mean at the case level? If you are doing process tracing, you are looking for sufficient conditions in specific cases. How does that generalise? The framework also clarifies why *equifinality* (multiple paths to the same outcome) is not a nuisance but a structural feature of social causation, and why we should expect causal heterogeneity across cases rather than treating it as an inconvenience.

Reading questions

Honours

1. In your own words, explain what an INUS cause is. Can you think of an example from a topic you are familiar with in political science: a factor that is part of a causal group but is neither individually sufficient nor strictly necessary?

2. Mahoney (2008) argues that case-oriented and population-oriented researchers are not studying different things, but are studying the same causal relationships at different levels. Does this claim strike you as persuasive? Why or why not?

MA/PhD

1. How does Mahoney's (2008) distinction between INUS and SUIN causes map onto the concepts of necessity and sufficiency? Under what conditions would identifying a cause as INUS rather than SUIN change the research design you would choose to study it?

2. Mahoney (2008) claims that mean causal effects are "symptoms" of INUS causes. What are the implications of this claim for how we interpret regression coefficients? If a coefficient captures the average effect of an INUS cause, what does that tell us (and not tell us) about causation in any individual case?

Morgan and Winship (2015), Ch. 1 and Ch. 2 (sections 2.1–2.2)

If Mahoney (2008) provides the conceptual map, Morgan and Winship (2015) provide the technical foundations. These two chapters introduce the potential outcomes framework (also called the Neyman-Rubin causal model) that underlies most contemporary quantitative causal inference. Chapter 1 is a broad overview that motivates the framework historically; the assigned sections of Chapter 2 formalise it. Together, they give you the notation and logic you need to understand what causal inference strategies are actually trying to achieve, which is what we will cover in Week 8.

Chapter 1: Why causal inference needed a framework

Morgan and Winship (2015) open with a historical sketch that is worth paying attention to, because it shows how the field arrived at the current framework and why earlier approaches were found wanting. The key story is the gradual recognition, from the 1930s onward, that observational data alone (no matter how much you have and no matter how sophisticated your statistical model) cannot resolve causal questions without additional assumptions about the data-generating process. The "age of regression," in which researchers assumed that controlling for enough variables in a regression model would isolate causal effects, gave way to a more cautious view in which the *design* of the study (not the complexity of the model) determines whether causal claims are credible. This is the "credibility revolution" in the social sciences, and it is the intellectual background for everything in Weeks 7 and 8.

The chapter walks through several applied examples including Catholic schools and educational achievement, school voucher lotteries, manpower training programmes, voting technology and ballot spoilage to illustrate the core problem. In each case, the question is the same: we observe that outcomes differ between groups (e.g., students who attend Catholic schools do better on tests than students who attend public schools), but we cannot tell whether the difference is caused by the group membership or by the pre-existing differences between the people who ended up in each group. This is the selection problem, and it is the reason the correlation-is-not-causation warning has teeth. It is not a vague philosophical worry; it is a concrete statistical problem with a specific structure.

Chapter 2 (2.1–2.2): The potential outcomes framework

Section 2.1 introduces the notation. For each unit i in a study, there are two *potential outcomes*: y^1_i (the outcome if unit i receives treatment) and y^0_i (the outcome if unit i does not receive treatment). The individual causal effect for unit i is the difference: $\delta_i = y^1_i - y^0_i$. The **fundamental problem of causal inference** is that we can only ever observe one of these two potential outcomes for any given unit. If a student attends a Catholic school, we observe y^1_i but not y^0_i; if they attend a public school, we observe y^0_i but not y^1_i. The other potential outcome

(the counterfactual) is missing data, and it is missing by definition, not by accident. No amount of data collection can solve this. Individual causal effects are therefore *unobservable*.

Connecting to Mahoney (2008)
Morgan and Winship's framework and Mahoney's framework are asking the same question from different directions. Morgan and Winship (2015) ask: across a population, what is the average difference in outcomes attributable to a treatment? Mahoney asks: in specific cases, was a factor part of a sufficient (or necessary) condition for the outcome? The potential outcomes framework gives us the tools to define causal effects precisely; Mahoney's (2008) framework gives us the tools to interpret what those effects mean at the case level. In your own research, you will likely lean on one more than the other, but understanding both makes you a more careful analyst regardless of method.

Reading questions

Both levels

3. Explain the fundamental problem of causal inference in your own words, using an example from political science (not one from the textbook). Why can we not simply compare people who experienced some event to people who did not?
4. Morgan and Winship (2015) argue that the "age of regression" led to overconfidence in causal claims. In your own experience reading political science, have you encountered studies that seem to treat statistical control as sufficient for causal inference? What would the potential outcomes framework say is missing from that approach?
5. Morgan and Winship distinguish between identification and statistical inference. In your own words, explain why a study could have a very large sample size and precisely estimated coefficients but still fail to identify a causal effect.

Cross-reading question
Morgan and Winship's (2015) potential outcomes framework defines causal effects at the population level as averages. Mahoney's framework (2008) defines causal effects at the case level in terms of necessity and sufficiency. Are these genuinely different definitions of what it means for X to cause Y, or are they different ways of looking at the same underlying causal reality? What is gained and what is lost by adopting each perspective?

Cho, Dreher, and Neumayer (2013)

This article serves as our applied example for the week. It asks a substantively provocative question (does legalising prostitution increase human trafficking?) and in doing so illustrates nearly every conceptual and inferential challenge we are discussing in this class. I have assigned it not because the answer to their research question is settled, but because the article is an unusually clear window into the difficulties of making causal claims from observational, cross-sectional data. Read it with a critical eye: the question is whether their research design supports the causal language they use.

The research question and theoretical framework
Cho, Dreher, and Neumayer (2013) argue that the legalisation of prostitution has two opposing effects on human trafficking inflows. The **scale effect** works through the expansion of the market: legalisation increases both demand (clients are no longer deterred by prosecution risk) and supply (sex workers are no longer deterred by the wage premium needed to compensate for illegality), so the equilibrium quantity of prostitution rises. If trafficked individuals

constitute a roughly constant share of prostitutes, then a larger market means more trafficking in absolute terms. The **substitution effect** works in the opposite direction: once prostitution is legal, businesses have incentives to recruit legal workers (domestic nationals or legal residents) rather than trafficked individuals, because employing trafficked persons now endangers their newly legal status. Theoretically, the net effect is indeterminate in that it depends on which effect dominates.

This is a clean illustration of the kind of theoretical ambiguity that motivates empirical research. The authors are transparent that theory alone cannot tell us the answer, which is commendable. Their empirical analysis uses cross-sectional data on up to 150 countries, with reported trafficking inflows (measured on an ordinal 0–5 scale from the UNODC (2006) Global Report on Trafficking in Persons) as the dependent variable and a dummy for whether prostitution is legal as the main independent variable.

This is where the article becomes most useful for our purposes. Read through the lens of the potential outcomes framework, the central question is whether countries that legalised prostitution would have experienced the same level of trafficking had they not legalised it. The *naive estimator* (comparing average trafficking levels between countries with and without legal prostitution) is exactly the kind of comparison Morgan and Winship (2015)warn us about. Countries do not randomly adopt prostitution laws. Countries that choose to legalise may differ systematically from countries that do not, in ways that also affect trafficking: wealthier countries, more open economies, more liberal political cultures, and countries in regions with high existing migration flows may be both more likely to legalise prostitution and more likely to experience trafficking inflows. This is the **selection problem**.

Cho, Dreher, and Neumayer (2013) are aware of this problem and attempt to address it by including control variables (GDP per capita, population, rule of law, democracy, migrant stock, share of Catholics, and regional dummies) and by using ordered probit estimation. They also conduct extensive robustness tests, including regional jackknife analysis and extreme bounds analysis. But (and this is the critical point for us this week) controlling for observables does *not* solve the fundamental identification problem. The concern is not just that the treatment and control groups differ on observable characteristics (which can, in principle, be controlled for statistically), but that they may differ on *unobservable* characteristics that also affect the outcome. The authors acknowledge this when noting that their cross-sectional design cannot control for unobserved country heterogeneity, and that they cannot find a valid instrumental variable for their independent variable. The regional fixed effects they include are a partial fix at best.

This article also illustrates the measurement issues we discussed in Weeks 3 and 4. The dependent variable is an ordinal index constructed from reports by 113 institutions, with all the problems that implies: reporting bias (wealthier countries with better institutions report more trafficking, which may reflect detection capacity rather than actual incidence), geographic bias (the sources over-represent Western Europe and North America), and the gap between reported and actual trafficking (since trafficking is clandestine by definition). The authors discuss these limitations candidly, but the fundamental concern remains: if the measurement of the dependent variable is correlated with the independent variable (countries that legalise prostitution may also be countries with better trafficking detection and reporting), then what looks like a causal effect of legalisation on trafficking may partly be a measurement artefact.

The authors supplement their cross-country analysis with brief case studies of Sweden, Germany, and Denmark, three countries that changed their prostitution laws in opposite directions during the study period. Sweden prohibited prostitution in 1999; Germany further legalised it in 2002; Denmark decriminalised self-employed prostitution in 1999. These case comparisons are interesting but also illustrate the limits of comparative case analysis when the number of cases is very small and there are many potential confounders. The Sweden-Denmark comparison is particularly useful: two countries with similar institutional and economic profiles that moved in opposite directions on prostitution law, with Denmark subsequently experiencing higher estimated trafficking. But even here, causal attribution is difficult: cross-border effects (Swedish clients travelling to Denmark), differences in enforcement, and the lack of reliable pre-reform trafficking data all complicate causal inference.

We are reading this article because it is an honest, methodologically serious attempt to answer a causal question with data that are not up to the task, which the authors are transparent about. It is an excellent example of the gap between the causal question a researcher wants to answer and the causal claims the available data and design can support. When you read next week's readings on experiments and quasi-experimental designs, think about what it would take to answer Cho, Dreher, and Neumayer's (2013) question more convincingly. What design would you need? What data? What assumptions?

Reading questions

Honours

6. Cho, Dreher, and Neumayer (2013) find that countries with legal prostitution have higher reported trafficking inflows. Does this mean that legalising prostitution causes more trafficking? What alternative explanations can you think of for this correlation?

7. The article identifies a "scale effect" and a "substitution effect" that work in opposite directions. In Mahoney's (2008) terms, what kind of cause is legalisation? Is it an INUS cause of trafficking?

MA/PhD

6. Using Morgan and Winship's (2015) potential outcomes framework, write out what the treatment, control, and potential outcomes are in Cho, Dreher, and Neumayer's (2013) study. What is the estimand they would ideally like to identify, and what assumptions would need to hold for their cross-sectional comparison to identify it?

7. The dependent variable is an ordinal index of *reported* trafficking inflows, not actual trafficking. If countries with legal prostitution are also countries with better detection and reporting of trafficking, how does this affect the interpretation of the main result? Which direction would this bias push the estimate?

8. Cho, Dreher, and Neumayer (2013) say that they cannot find a valid instrument for their prostitution variable. What would a good instrument need to satisfy (relevance and the exclusion restriction)? Can you think of any plausible candidate? Why is finding one so difficult in this context?

PART 3: GROUP ACTIVITY

Diagnosing causal claims in published research
This activity asks you to take the conceptual tools from today's readings and apply them to the kinds of causal claims you encounter in published political science research. The goal is to develop the habit of reading causal claims critically, not cynically, but precisely: asking *what assumptions does this claim require, and are those assumptions plausible?*

Your task
Working in your groups, you will each receive a brief extract from a published political science study that makes a causal claim (I will distribute these in class). For your assigned extract, work through the following diagnostic questions together. You have about 20 minutes, then each group will briefly present their diagnosis to the class.

Step 1: Identify the causal claim

What does the study claim causes what? Write the claim in the form "X causes Y." Identify the treatment (X), the outcome (Y), and the units of analysis. Is the claim about an average effect across a population, or about causation in specific cases, or both?

Step 2: Apply the potential outcomes framework

What are the potential outcomes? For a unit in the study, what is y^1 (outcome under treatment) and y^0 (outcome under control)? What is the counterfactual that the study needs to construct? Does the research design provide a credible approximation of this counterfactual, or is the comparison group likely to differ from the treatment group in ways that would bias the estimate?

Step 3: Apply Mahoney's (2008) framework
Thinking about the treatment variable X: is it best understood as an INUS cause, a SUIN cause, a necessary condition, or a sufficient condition? Is the study's design suited to the type of cause it is? (For example, if X is plausibly an INUS cause, one component of a larger causal group, does the study account for the other components of the group? If the study finds a mean causal effect, what would Mahoney (2008) say that effect tells us about causation in individual cases?)

Step 4: Identify the key threats
What are the most important threats to the causal interpretation of this study's findings? Think specifically about: confounders (what unobserved variables could produce both X and Y?), reverse causation (could Y cause X instead?), selection bias (are the units that received the treatment systematically different from those that did not?), and measurement issues (is the treatment or outcome measured in a way that could produce spurious associations?). Which single threat do you consider most serious, and why?

Step 5: Suggest a stronger design
If you could redesign this study with unlimited resources, what would you do differently to strengthen the causal claim? Think about what kind of variation in the treatment you would need, what comparison group would be most credible, and what additional data would help. You do not need to be technically specific about methods (we will cover some of those next

week), but you should be able to articulate what a more convincing counterfactual would look like.

Each group will have about 3 minutes to present their diagnosis. Focus on what the causal claim is, what the key threat is, and what a stronger design would look like. Expect questions from other groups about whether you have identified the most important threat.

PART 4: WRAPPING UP AND LOOKING AHEAD

Let me briefly pull together the main threads from today. First, causal inference is hard for a specific reason: the fundamental problem of causal inference means we can never directly observe causal effects. Every causal claim requires us to construct or argue for a counterfactual and different research designs are different strategies for making that counterfactual credible. Second, causation is not just about average effects. Mahoney's (2008) framework reminds us that causes can be parts of causal groups (INUS causes), that the same outcome can be produced by different causal paths (equifinality), and that the population-level effects we estimate with regression or experiment are connected to, but not the same as, the case-level causal processes that qualitative researchers study. Understanding both levels makes you a better designer and a better consumer of research. Third, the Cho, Dreher, and Neumayer (2013) article illustrates what happens when you have a good causal question but a research design that cannot fully answer it. Their work is methodologically careful and transparent about its limitations, but the fundamental identification problem remains that cross-sectional observational data with no credible instrument or source of exogenous variation cannot establish that legalised prostitution *causes* more trafficking, even if the correlation is robust. The design problem is not laziness; it is the nature of the question and the available data.

Next week we turn from the problem to the solutions. Week 8 will cover the major research design strategies for credible causal inference: randomised experiments, natural experiments, instrumental variables, difference-in-differences, and regression discontinuity. Each of these is a different way of generating the ignorability condition (or something close to it) that we established today is necessary for unbiased causal estimates. The conceptual groundwork we are laying this week is essential. You need to understand why causal inference is hard before you can appreciate what these designs are doing and why they work (and when they don't).