Research Design in Political Science
POLS4011/POLS8058

Richard Frank
31 March 2026

WEEK 6: CASE SELECTION AND SCOPE, PART 2

PART 1: OVERVIEW

This is the second week on case selection and scope. Last week focused on the logic of case selection: *which* cases to choose, *how many*, and *why*. Importantly, Seawright and Gerring (2008) gave us a typology of case selection techniques. Geddes (1990) showed what goes wrong when you select on the dependent variable, and Tannenwald (1999) showed us the theoretical and empirical power of a well-chosen case. This week we extend that discussion into scope conditions, external validity, and the question of what you can claim about the world beyond the cases you studied.

The shift from Week 5 to Week 6 is a shift in emphasis. Last week we asked: how do I choose my cases well? This week we ask: once I have chosen my cases and produced findings, what can I legitimately claim those findings apply to? This is the question of external validity, and it is where many research projects become either overconfident or modest. A study with strong internal validity (the causal inference is sound within the cases studied) may have weak external validity (the findings do not travel well to other contexts). A study that maximises generalisability may sacrifice the depth needed for credible causal claims. This tension between internal and external validity is not a problem to be solved but a trade-off to be managed, and managing it well requires being explicit about your scope conditions.

Scope conditions are the boundaries you draw around your argument. They specify the conditions under which you expect your theory to hold and, equally importantly, the conditions under which you do not expect it to hold. Every theory has scope conditions, whether the author states them or not. The difference between a well-designed study and a less than well-designed one is often that the former is explicit about its scope while the latter lets the reader guess its scope. This connects to our conceptual discussions in Weeks 3-4. If your concept is defined in a way that is culturally or temporally specific (recall Collier and Levitsky's [1997] diminished subtypes, or the measurement validity challenges in Munck and Verkuilen [2002]), then your scope conditions need to reflect that specificity. A finding about competitive authoritarianism in post-Soviet states may not apply to competitive authoritarianism in sub-Saharan Africa in 2026 if the concept is operationalised differently in the two contexts.

This week's readings approach scope and external validity from three different angles. Collier and Mahoney (1996) continue the selection bias debate from last week but complicate it considerably (sigh). They argue that the standard warnings about selection bias, which originate in econometrics, do not translate straightforwardly to qualitative research. The frame of comparison matters. Whether selecting on the dependent variable introduces bias depends on what question you are asking, how you define the dependent variable, and what comparison you have in mind. This is a more nuanced treatment than Geddes (1990), and it pushes back against the blanket prohibition on selecting cases with extreme outcomes. Lieberman (2005) offers a practical framework for combining large-N statistical analysis with small-N case studies in what he calls nested analysis. His article addresses the scope question by showing how the two methods can inform and discipline each other, with the large-N analysis guiding case selection for the small-N component and the small-N analysis testing the robustness of the statistical findings. Fortna (2004) is our applied example of peacekeeping effectiveness after civil wars that directly confronts a classic selection bias problem (peacekeepers are sent to the hardest cases) and uses quantitative methods to address it.

It is worth thinking about Hyde (2007) again this week. Her Armenian election study has been a recurring example throughout the course, and this week it helps illustrate the scope question. Hyde's (2007) natural experiment provides strong internal validity: within the Armenian election, the quasi-random assignment of monitors allows a credible causal estimate of monitoring effects on fraud. But what is the external validity of that finding? Does it generalise to elections in other post-Soviet states? To elections in Sub-Saharan Africa? To elections where the incumbent is not strategically allowing monitors? Hyde's (2007) later work has explored these scope questions, and it may be useful for you to think about how the case selection and scope tools from Weeks 5 and 6 would help you evaluate the boundaries of her finding.

Plan for today

1. Overview: connecting Week 5's case selection logic to Week 6's scope and external validity questions
2. Readings: selection bias in qualitative research, nested analysis, and controlling for selection effects in quantitative research
3. Group activity
4. Critical review check-in and looking ahead to causal inference

Key topics this week

- Selection bias in qualitative research is more nuanced than the standard econometric critique suggests. The appropriate frame of comparison matters.
- Within-case analysis (process tracing) can partially address selection bias, but it cannot fully overcome it.
- Mixed methods are not just about combining qualitative and quantitative work. The combination needs to be structured so that each method disciplines the other.
- Scope conditions are not an afterthought. They are a core design decision that determines what your findings mean beyond your cases.
- The tension between internal and external validity is a trade-off to be managed, not a problem to be eliminated.
- Addressing selection bias in quantitative research requires thinking carefully about what determines both the treatment and the outcome.

The differentiated expectations continue. Honours students should be able to ***identify the scope conditions*** of their own project and ***explain*** what their findings would and would not generalise to. MA/PhD students should be able to ***evaluate the trade-offs*** between internal and external validity in their design, ***defend their scope choices*** against plausible alternatives, and ***articulate how mixed-method strategies*** could strengthen (or would not strengthen) their inferential claims.

PART 2: READINGS

1. Collier, David and James Mahoney 1996. "Insights and Pitfalls: Selection Bias in Qualitative Research." *World Politics* 49(1): 56–91.

2. Lieberman, Evan S. 2005. "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99(3): 435–452.

3. Fortna, Virginia Page. 2004. "Does Peacekeeping Keep Peace? International Intervention and the Duration of Peace After Civil War." *International Studies Quarterly* 48(2): 269–292.

1. Collier and Mahoney (1996)

Collier and Mahoney (1996) is arguably the most abstract reading this week, but it is worth the effort. While Geddes (1990) offers a clear and direct warning about selection on the dependent variable, Collier and Mahoney (1996) complicate her warning in some interesting ways. Their central argument is that the concept of selection bias, which originates in econometrics and is defined precisely in that context, does not transfer to qualitative research without significant modification. The standard critique (selecting cases with extreme values on the dependent variable biases causal inferences) is correct in the quantitative context from which it comes. But qualitative researchers ask different questions, use different methods of causal assessment, and operate within different frames of comparison, all of which affect whether and how selection bias actually operates.

Collier and Mahoney (1996) begin with the standard quant argument. If you truncate your sample by selecting only cases with high (or low) values on the dependent variable, you systematically underestimate the strength of the causal relationship. Collier and Mahoney (1996) illustrate this with a figure showing how restricting the range of the dependent variable flattens the regression slope. Importantly, they note an asymmetry: selection on the *dependent* variable biases estimates, but selection on the *explanatory* variable does not (in the bivariate case). This distinction matters because qualitative researchers sometimes select cases based on their values on the explanatory variable (which is fine) but are criticised as though they have selected on the dependent variable (which would be problematic).

The concept of the "frame of comparison" or "contrast space" is central here. Collier and Mahoney (1996) argue that whether selection bias is present depends on what comparison the researcher intends. If a scholar studying high-performing governments restricts their analysis to cases with very good outcomes, this is only biasing if they intend to make claims about the full range of government performance. If their question is specifically about what distinguishes the best performers from each other, the narrower frame of comparison may be appropriate. The key insight is that selection bias is not an inherent property of a set of cases. It is a relationship between the cases chosen and the inferential claim being made. A study that would be biased for one research question might not be biased for another, even with the same cases. Clearly a more refined if harder to implement approach than that spelled out in Geddes (1990).

Collier and Mahoney (1996) also introduce the problem of causal heterogeneity. This is the possibility that the causal process producing extreme outcomes differs from the causal process operating in the broader population. If causal heterogeneity is present, then the broader comparison that the standard selection bias critique demands may itself be misleading, because it assumes causal homogeneity across all cases. For qualitative researchers who are often sceptical that the same causal process operates everywhere, this is an important point. It suggests that narrowing the frame of comparison to a more homogeneous set of cases may sometimes be more appropriate than broadening it, even though broadening is what the standard selection bias advice (e.g., KKV 1994) recommends.

A particularly useful section of the article given our Week 5 in-class discussion relates to whether within-case analysis can overcome selection bias. Collier and Mahoney (1996) are cautiously sceptical. Within-case analysis (process tracing, pattern matching, causal narrative) is a powerful tool for establishing causal connections within a single case, and it can help identify causal mechanisms that cross-case analysis cannot detect. However, it cannot, on its own, solve the selection bias problem, because selection bias is fundamentally about what the cases you study can tell you about cases you did not study. Process tracing within a truncated sample can show you how a causal process worked in the cases you examined, but it cannot tell you whether that same process operates in the cases you excluded. This connects to our earlier discussion of internal versus external validity: within-case analysis strengthens internal validity, but it does not, by itself, guarantee external validity.

Finally, Collier and Mahoney (1996) raises the problem of "complexification": the tendency for case study researchers who examine extreme cases to discover additional causal factors that appear important in those cases but may not generalise. When you look closely at extreme outcomes, you often find complexity. That complexity may be real and important, or it may be an artefact of having selected unusual cases where many things came together in unusual ways. Collier and Mahoney (1996) do not provide a clear answer to complexification, but they flag it as a distinctive risk of case studies that select on extreme outcomes.

This article is interesting to compare to Geddes (1990) from last week. Geddes (1990) gives us the clear, sharp version of the selection bias critique. Collier and Mahoney (1996) give us a more nuanced, complicated version that directly critiques Geddes (1990). Both are incredibly influential and useful to understand, but they operate at different levels of complexity. For your critical reviews and final research design papers, you need to be able to deploy both the basic logic of why selection on the dependent variable is problematic, and the more sophisticated understanding of when that critique applies more or less forcefully depending on your research question, the frame of comparison, and the method of analysis.

Reading questions

Honours students

1. Collier and Mahoney (1996) argue that whether selection bias is present depends on the frame of comparison. In your own words, explain what this means. Why does the same set of cases produce bias for one research question but not another?

2. Collier and Mahoney (1996) distinguish between selection on the dependent variable and selection on the explanatory variable. Why does this distinction matter? For your own project, which variable are you selecting on, and what are the implications?

MA/PhD students

1. Collier and Mahoney (1996) suggest that causal heterogeneity can justify a narrower frame of comparison, even though the standard selection bias critique recommends broadening it. Evaluate this argument. Under what conditions is narrowing the comparison defensible, and when does it become a rationalisation for convenient case selection?

2. The article argues that within-case analysis cannot fully overcome selection bias. If this is correct, what are the implications for single-case research designs? Is there a way to combine within-case and cross-case methods that addresses both internal and external validity, or is there always a residual trade-off?

2. Lieberman (2005)

Lieberman (2005) offers a practical and quite systematic framework for mixed-method research. He addresses a problem that many of us face in our own work: we have access to both quantitative data and qualitative knowledge (case expertise, process-tracing evidence, archival material), but we are not sure how to combine them in a way that is more than just doing two separate studies stapled together. Lieberman's (2005) answer is nested analysis, a structured approach in which large-N analysis (LNA) and small-N analysis (SNA) inform and discipline each other within a single research design. In my experience, multi-methods honours, MA, and PhD theses have become more common over time, but the justification for this approach is no where near as clear and compelling as what is outlined here.

Lieberman's (2005) framework begins with a preliminary quant analysis. The LNA serves several purposes: (1) it identifies the range of variation on the dependent variable, (2) it estimates the strength and direction of relationships between variables, and (3) it provides a baseline against which individual cases can be evaluated. Importantly, the LNA also helps guide case selection for the small-N component. This is where Lieberman's (2005) framework connects to Seawright and Gerring's (2008) case selection typology from last week and to the scope questions we are discussing this week.

After the preliminary LNA, the researcher assesses whether the results are robust and satisfactory. If they are, the researcher proceeds to what Lieberman (2005) calls Model-testing Small-N Analysis (Mt-SNA). Here, the goal is to use case studies to test the robustness of the statistical findings. Cases should be selected that are "on the line" (cases that are well predicted by the statistical model, what we referred to last week as "typical" cases). The logic is that if the model correctly predicts these cases, process tracing within them should reveal the causal mechanisms the model implies. If the process tracing contradicts the model's implications, this is evidence that the statistical relationship may be spurious or that the causal mechanism is different from what the model assumes. For Mt-SNA, cases can be selected either deliberately (on the line) or randomly from among well-predicted cases.

If the preliminary LNA results are *not* robust or satisfactory, the researcher takes a different path: Model-building Small-N Analysis (Mb-SNA). Here, the goal is to use case studies to develop a better theoretical model. Cases should be selected that are "off the line" (poorly predicted by the statistical model). The logic is that these are the cases where the model fails, and studying them closely may reveal omitted variables, alternative causal pathways, or scope conditions that the model does not capture. Mb-SNA is essentially theory-generating rather than theory-testing, and it connects to the deviant case logic in Seawright and Gerring (2008) last week.

Lieberman's (2005) framework is useful because it makes the relationship between quantitative and qualitative components *sequential and structured* rather than (as it often is) ad hoc. The LNA informs the SNA (by guiding case selection), and the SNA feeds back into the LNA (by testing or refining the model). This iterative structure means that each component adds inferential value that the other cannot provide on its own. The LNA provides external validity that the SNA lacks; the SNA provides causal depth and mechanism identification that the LNA lacks. Together, they address the internal/external validity trade-off more effectively than either method alone.

For the scope questions we are discussing this week, Lieberman (2005) provides an important lesson: scope conditions are not just something you state at the end of a study. They are something your research design should actively explore. If your LNA identifies cases where the model fails, those failures tell you something about where your theory's scope ends. If your SNA reveals that the causal mechanism works differently in different contexts, that tells you something about the conditions under which your argument holds. A well-executed nested analysis does not just produce findings; it produces findings with built-in information about their own boundaries.

I want to stress that not every project can (or should) use nested analysis. It requires access to both quantitative data and case-study material, it assumes the LNA has enough cases for meaningful statistical analysis (Lieberman suggests at least 12 for cross-national work, which I think is still quite low), and it requires us to have substantial quant and qual skills. But even if you do not use nested analysis in your own research, the logic is useful. It shows how case selection can be guided by systematic criteria rather than convenience, and it demonstrates that the choice between qualitative and quantitative methods is often a false dichotomy. The real question is how to combine methods in a way that each compensates for the other's weaknesses.

Reading questions

Honours students

3. Lieberman (2005) distinguishes between on-the-line and off-the-line case selection. In your own words, explain the difference and when you would use each. Can you identify where your own case would fall relative to a hypothetical regression line?

4. Lieberman (2005) argues that combining large-N and small-N analysis produces stronger inferences than either alone. What does each component contribute that the other cannot? Think about a version of your own project that used both methods. What would the large-N component look like, and what would the case study component add?

MA/PhD students

3. Lieberman (2005) assumes that a preliminary LNA can meaningfully guide case selection. But what if the available quantitative data are poorly measured or the model is mis-specified? How robust is the nested analysis framework to weaknesses in the LNA component, and what safeguards does Lieberman offer against being misled by a flawed statistical baseline?

4. Compare Lieberman's (2005) nested analysis framework with the case selection advice in Seawright and Gerring (2008). Where do they agree, and where do they diverge? Does nested analysis resolve or merely repackage the fundamental tension between internal and external validity?

3. Fortna (2004)

Fortna (2004) is this week's applied example, and she confronts head-on a selection bias problem that many applied researchers face: the treatment you want to evaluate is not randomly assigned. Her research question is whether peacekeeping after civil wars helps maintain peace, but the challenge is that peacekeepers (unlike election observers in Armenia) are not deployed randomly. The international community tends to send peacekeepers to the hardest cases, conflicts where peace is most fragile and the risk of renewed fighting is highest. If you simply compare cases with peacekeepers to cases without them, you are likely to underestimate the effect of peacekeeping, because the peacekeeping cases started with worse baseline conditions. This is a textbook selection bias problem, and it is the quantitative analogue of the issues Collier and Mahoney (1996) discuss in the qualitative context.

Fortna's (2004) strategy for addressing this problem has two parts. First, she models *where* peacekeepers get sent. Using logit models, she identifies the factors that predict peacekeeping deployment: wars that end without a decisive military victory, conflicts in states with small armies, and (less consistently) wars with higher death tolls are all more likely to receive peacekeepers. This analysis is not the main research question, but it is essential for addressing selection bias. By understanding what determines the treatment (peacekeeping), Fortna (2004) can control for those factors when estimating the treatment's effect. This is the logic behind what quantitative methodologists call controlling for confounders (variables that affect both the likelihood of treatment and the outcome).

Second, Fortna (2004) uses a Cox proportional hazards model (a type of duration model) to estimate the effect of peacekeeping on the duration of peace after civil war. Duration models are appropriate here because the outcome of interest is not simply whether peace holds or breaks down, but *how long* peace lasts. The Cox model estimates hazard ratios: a ratio below 1 means the variable is associated with longer peace (lower risk of war resuming), while a ratio above 1 means shorter peace (higher risk). By including the factors that predict peacekeeping deployment as control variables, Fortna (2004) attempts to compare peacekeeping and non-peacekeeping cases that are similar in their baseline degree of difficulty. The analogy she uses is suitable: just as a study of medical treatment effectiveness must control for how sick the patients were before treatment, a study of peacekeeping must control for how difficult the conflict environment was before deployment.

The substantive findings are worth discussing for their methodological lessons. The crosstab in Table 1 (p. 272) suggests that peacekeeping makes little difference or may even be associated with *worse* outcomes. However, this is exactly what selection bias looks like: peacekeepers go to hard cases, so the direct comparison is misleading. Once Fortna (2004) controls for the degree of difficulty, peacekeeping after the Cold War reduces the risk of renewed war by ~70%. The difference between the simple comparison and the estimated one is a concrete illustration of how important selection bias can be, and why the methodological issues we have been discussing are not abstract concerns but practical ones that affect substantive conclusions.

Fortna (2004) also distinguishes between four different types of peacekeeping missions: observer missions, traditional peacekeeping, multidimensional peacekeeping, and peace enforcement. This is itself a conceptual exercise that connects back to Weeks 3-4. If you lump all peacekeeping together, you may miss important variation. Observer missions (small, unarmed) may work differently from enforcement missions (large, armed, mandated to use force). By disaggregating the treatment, Fortna (2004) can ask more precise questions about what kinds of peacekeeping work and under what conditions, which is a scope question.

For this class, I think the main lessons from Fortna (2004) are threefold. First, selection bias is not only a qualitative research problem. It arises whenever the treatment of interest is systematically related to the outcome, which is common in observational research of all kinds. Second, addressing selection bias requires understanding what determines the treatment, not just what determines the outcome. Fortna (2004) had to model peacekeeping deployment before she could credibly estimate its effects. Third, the bivariate analysis and the controlled analysis can lead to opposite conclusions. This is a clear demonstration of why the design issues we have been discussing all term matter for substantive findings, not just for methodological neatness.

Reading questions

Honours students

5. Fortna (2004) argues that the raw comparison between peacekeeping and non-peacekeeping cases is misleading because of selection bias. In your own words, explain why. What would happen if a researcher drew conclusions from the raw comparison without controlling for the degree of difficulty?

6. Fortna (2004) disaggregates peacekeeping into different mission types. How does this connect to the conceptual discussions in Weeks 3 and 4? What would be lost if she had treated all peacekeeping missions as a single category?

MA/PhD students

5. Fortna's (2004) strategy for addressing selection bias relies on controlling for observable confounders. What are the limitations of this approach? Can you think of unobservable factors that might affect both peacekeeping deployment and peace duration that her controls would miss? How would this affect her conclusions?

6. Compare Fortna's (2004) approach to selection bias with the approaches discussed in Collier and Mahoney (1996) and Lieberman (2005). Fortna uses quantitative controls; Collier and Mahoney(1996) discuss reframing the comparison; Lieberman (2005) proposes combining methods. Are these complementary or competing strategies? Under what conditions would each be most appropriate?

Overall reading questions

1. Collier and Mahoney (1996) argue that the standard selection bias critique needs to be adapted for qualitative research. Geddes (1990) presented the critique more directly. Are these two positions contradictory, or can they be reconciled? What would you need to understand from both readings to make good case selection decisions in your own project?

2. Lieberman (2005) proposes that large-N analysis should guide case selection for small-N analysis. Collier and Mahoney (1996) suggest that the appropriate frame of comparison depends on the research question and may sometimes be narrow. When would Lieberman's (2005) approach and Collier and Mahoney's (1996) approach give you different advice about how to select cases, and how would you adjudicate between them?

3. Think again about Hyde (2007). Her Armenian election study has strong internal validity from the natural experiment, but its external validity is a separate question. Using this week's readings, how would you evaluate the scope conditions of Hyde's (2007) finding? What would Lieberman's (2005) nested analysis framework suggest as a next step for extending or testing her finding in other contexts?

4. In the last two weeks we have covered case selection strategies (Seawright and Gerring 2008), the dangers of selecting on the dependent variable (Geddes 1990), crucial case logic (Tannenwald 1999), the nuances of selection bias in qualitative research (Collier and Mahoney 1996), a framework for mixed-method research (Lieberman 2005), and a quantitative approach to addressing selection effects (Fortna 2004). Synthesise what you have learned. If you were advising a fellow student rwho is just beginning their research design, what are the three most important things they need to understand about case selection and scope?

PART 3: GROUP ACTIVITY

This activity has you apply Lieberman's (2005) nested analysis framework to Fortna (2004). We are taking a break from thinking about your own research and writing to design a model-testing small-N analysis (Mt-SNA) that could follow from Fortna's (2004) large-N analysis. This requires you to engage with both readings simultaneously. You need to understand what Fortna (2004) found, and you need to apply Lieberman's (2005) logic to decide how you could test those findings with case studies.

Background: Fortna's (2004) key findings
Fortna's (2004) preliminary LNA (in Lieberman's [2005] terms) uses Cox proportional hazards models to estimate the effect of peacekeeping on the duration of peace after civil wars. Her core finding is that in the post-Cold War period, the presence of peacekeepers reduces the risk of renewed war by approximately 70% (Table 7, p. 274, hazard ratio of 0.32 for all peacekeeping). She also finds that (1) consent-based missions (observer, traditional, and multidimensional peacekeeping) are more effective than enforcement missions, that (2) peace is more stable after decisive military victories, and that (3) higher war costs are associated with shorter peace. Crucially, the simple bivariate comparison in Table 1 (p.272) is misleading. Peacekeeping appears to make little difference until you control for the fact that peacekeepers are sent to the hardest cases.

Your mission
Working in your groups, propose a model-testing small-N analysis that would follow from Fortna's (2004) LNA. Work through the following steps together. You have about 20 minutes, then each group will briefly present their proposed design to the class.

Step 1: Assess the LNA
Lieberman (2005) says the first decision in nested analysis is whether the preliminary LNA results are "robust and satisfactory." Consider Fortna's (2004) results: the core peacekeeping finding is statistically significant and substantively large in the post-Cold War period, but several control variables (treaty, identity war, factions, democracy) are not consistently significant, and the sample is relatively small (N = 52 for the post-Cold War period). On balance, would you assess these results as robust enough to proceed with model-testing SNA, or would you lean toward model-building SNA? Justify your assessment.

Step 2: Select your cases

Assuming you proceed with Mt-SNA, Lieberman (2005) advises selecting cases that are "on the line," that is, cases well predicted by the statistical model. Using Fortna's (2004) Table 3 (p. 274), which lists all post-Cold War cases by peacekeeping status and outcome, select one to two specific cases for in-depth analysis. You should choose cases where the outcome (peace held or war resumed) is consistent with what Fortna's model would predict given the case's characteristics. For example, a case with peacekeepers where peace held, or a case without peacekeepers that ended in a decisive victory and where peace also held. Consider selecting cases that vary on the key independent variable (peacekeeping presence) so you can trace the causal mechanism in contrasting contexts. Be specific. Name the cases and explain why each is "on the line."

Step 3: Design the small-N analysis
For the cases you selected, what causal mechanisms would you look for through process tracing? Fortna's (2004) statistical model tells us that peacekeeping is associated with longer peace, but it does not tell us how peacekeeping produces that effect. What specific observable evidence would confirm that peacekeeping actually caused peace to last in your chosen cases? Think about what you would need to find in the historical record, in interview data, or in archival sources to demonstrate that the statistical relationship reflects a genuine causal process rather than a spurious correlation.

Step 4: Anticipate what you might find
Lieberman (2005) stresses that Mt-SNA must plan for the possibility that the case study does not support the model. If your process tracing reveals that peacekeeping was not actually the reason peace held in your "on-the-line" cases (perhaps peace held for idiosyncratic reasons, or the causal mechanism was different from what you expected) what would that imply for Fortna's (2004) findings? Would you conclude that the case was idiosyncratic (and the model is still sound), that there are theoretical flaws in the model, or that you need to shift to model-building SNA? Refer to the decision tree in Figure 1 of Lieberman (2005: 437) to explain where in the nested analysis process you would go next.

Each group will have about three minutes to present their proposed Mt-SNA design. Focus on which cases you chose, why they are "on the line," what causal evidence you would look for, and what you would do if the case study did not support the model. Hint: expect questions from other groups about your case selection choices. This is where Seawright and Gerring's (2008) typology should be helpful.

PART 4: CRITICAL REVIEWS AND LOOKING AHEAD

Hopefully, you are well advanced in your critical review writing. I want to spend a bit of time today answering any final questions and dealing with any potential ambiguities in, or confusion about my critical review guide

Looking ahead. After the break Weeks 7 and 8 turn to causal inference. The first six weeks have been building toward this: concepts (what are we measuring?), case selection (where are we measuring it?), and scope (what can we claim beyond what we measured?). The next two weeks will ask the question that motivates most empirical research in political science: can we establish that X caused Y? We will discuss the logic of causal inference, the potential outcomes framework, and the tools researchers use to make causal claims credible including experiments, natural experiments, instrumental variables, difference-in-differences, and regression discontinuity. The case selection and scope discussions from this week and last week are relevant, because every causal inference strategy involves assumptions about which cases are comparable and what the findings generalise to.