Research Design in Political Science
POLS4011/POLS8058

Richard Frank
10 March 2026

WEEK 3: CONCEPTS UNDER PRESSURE, PART 1

PART 1: OVERVIEW

This is the first of a two-week block on concepts and measurement. This week focuses on *concept formation*: what concepts are, how they are structured, and what goes wrong when they are poorly formed. Next week will turn to *operationalisation and measurement*: how concepts are translated into variables and indicators, and what validity and reliability demand. The division is deliberate: you cannot measure well what you have not defined well. Many measurement problems in political science are really conceptual problems in disguise.

The research design memo is due this Friday. This is not a coincidence. Every student submitting a memo is making initial conceptual commitments. You are naming the things you plan to study. This week's readings provide tools for evaluating whether those commitments are well-formed: whether the concepts are clearly defined, internally coherent, and distinguishable from neighbouring concepts. You should be reading this week's material with one eye on your own project and asking: *do I actually know what I mean by the key terms in my research question?*

"Concepts under pressure" is meant to capture a specific problem. In the abstract, many political science concepts seem clear enough (e.g., democracy, corruption, state capacity, and political violence). But the moment you try to do empirical work with these concepts, they come under pressure. You have to decide exactly what counts and what does not, where the boundaries are, and whether your definition is consistent with how others in the field use the term. This is where conceptual clarity becomes a research design issue, not merely a philosophical one. A poorly formed concept does not just produce bad definitions; it produces bad research designs, because the researcher does not know what they are looking for, cannot tell when they have found it, and cannot communicate their findings to others.

Plan for today

1. Overview: connecting Week 2 to Weeks 3 & 4
2. Readings: what concepts are, how they are structured, and what goes wrong
3. Group activity: putting your own concepts under pressure

Key themes for this week

- The difference between a word and a concept: why naming something is not the same as defining it.
- How concepts are structured: dimensions, levels, and the logical relationships between them.
- Conceptual stretching: what happens when a concept is applied beyond the domain where it was developed.
- The ladder of abstraction: trading off between generality and precision

- Why concept formation is a research design choice, not merely a definitional exercise
- The relationship between how you define a concept and what cases count as instances of it.

The differentiated expectations continue. Honours students should be able to ***identify*** the core concepts in their project and explain ***what*** they mean by them. MA/PhD students should be able to evaluate ***alternative*** definitions of their core concepts, articulate why they have chosen one over another, explain what their choice includes and excludes, and explain what the consequences are for their design.

PART 2: READINGS

Required readings

1. Adcock, Robert and David Collier (2001), "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research," *American Political Science Review* 95(3): 529–546.
2. Goertz, Gary (2006), *Social Science Concepts: A User's Guide*, Chapter 1.
3. Collier, David and Steven Levitsky (1997), "Democracy with Adjectives: Conceptual Innovation in Comparative Research," *World Politics* 49(3): 430–451.

1. Adcock and Collier (2001)

This article introduces a popular framework for thinking about how concepts relate to measures. Its main contribution is a four-level model of the concept–measurement relationship, combined with a typology of validity assessments that apply at each level. The article is explicitly pluralist: Adcock and Collier (2001) argue that the framework applies to quantitative and qualitative work, and that disputes about measurement validity are often really disputes about which level of the framework scholars are working at.

The four levels
The framework distinguishes four levels in the process of moving from an idea to a number (or a classification). The *background concept* is the broad constellation of meanings and understandings associated with a given concept as used in ordinary and academic discourse: what people in the field generally mean when they use a term like "democracy" or "corruption." The *systematised concept* is the specific formulation a researcher adopts for a given study: a deliberate selection from and refinement of the background concept. *Indicators* are the observable phenomena used to operationalise the systematised concept: the things you look for in the world. And *scores* are the values assigned to cases based on those indicators. Each transition between levels involves choices, and each choice can be evaluated for validity.

Why disputes arise
What I found interesting about this article is their claim that many our disagreements about whether something has been "properly measured" are disagreements occurring at different levels. When one academic says another's measure of democracy is invalid, the disagreement might be about the background concept (they mean fundamentally different things by "democracy"), or the systematised concept (they agree on the background concept but the operational definition is too narrow or too broad), or the indicators (the definition is fine, but the chosen indicators do not capture it), or the scoring rules (the indicators are fine but the

coding is unreliable). Clarifying *where* the disagreement sits is very useful. It turns vague complaints about measurement into addressable challenges.

Three types of validity assessment

Adcock and Collier (2001) identify three types of validity relevant to the concept–measurement link. *Content validity* (also called face validity or logical validity) asks whether the systematised concept and its indicators adequately capture the full content of the background concept. Has anything important has been left out or anything unnecessary has been included? *Convergent/discriminant validity* asks whether a given indicator or score is associated with (or distinguished from) alternative indicators or scores that are, in theory, related to (or distinct from) the same concept. *Nomological/construct validity* asks whether the scores produced by a measure relate to other variables in the way that the underlying theory predicts. Does the concept behaves the way it should within a broader causal or correlational framework? The first type is most relevant for this week; the latter two will become more important in Week 4 when we turn to measurement.

Connecting to last week

Adcock and Collier's (2001) framework connects to several things from last week. Van Evera's (1997) emphasis on clearly specified variables assumes that the concepts behind those variables have been already systematised. Booth et al.'s (2016) three-step formula asks students to name what they are studying, but naming is not defining. Adcock and Collier (2001) push us to go further. Once you have named your key concept (Booth et al.'s 2016 step 1), you need to systematise it (what exactly do you mean?), choose indicators (what will you look for?), and develop scoring rules (how will you assign values?). Those of you whose research design memos use terms like "democratic quality" or "political instability" without specifying which of many possible definitions they intend are working with unsystematised concepts, and their research designs will also be vague.

2. Goertz (2006), Chapter 1

Where Adcock and Collier (2001) provide a framework for the *process* of moving from concepts to measures, Goertz (2006) provides a framework for the *internal structure* of concepts themselves. His central argument is that concepts in the social sciences have a three-level architecture, and that the logical relationships between these levels have direct consequences for how concepts are operationalised, how cases are classified, and ultimately what findings a study can produce.

The three-level structure

Goertz proposes that concepts have a *basic level* (the concept itself, the word we use… i.e., "democracy," "welfare state," or "corruption"), a *secondary level* (the dimensions of the concept, i.e., the components that make it what it is), and an *indicator/data level* (the empirical measures used to assess whether and to what degree the concept is present). This parallels Adcock and Collier's (2001) framework but adds an important question: *what is the logical relationship between the secondary-level dimensions?*

The structure question: AND versus OR

When a concept has multiple dimensions, there are two ways those dimensions can relate to the overall concept. In an *essentialist* (or necessary/sufficient) structure, *all* dimensions must be present for the concept to apply. The dimensions are joined by "AND." If one dimension is absent, the case is not an instance of the concept. Democracy might be defined this way:

competitive elections AND civil liberties AND rule of law. Remove any dimension, and it is no longer a democracy. In a *family resemblance* structure, dimensions are joined by "OR." A case qualifies as an instance of the concept if it has *enough* of the dimensions, even if some are absent. A welfare state might be defined this way: generous pensions OR universal healthcare OR unemployment insurance OR public housing. A country does not need to have all four to count.

Why the structure matters for research design

The choice between AND and OR has important practical consequences. Under a necessary and sufficient structure, a case that scores zero on any single dimension is automatically classified as zero, regardless of how well it scores on other dimensions. Under a family resemblance structure, weakness on one dimension can be compensated by strength on another. These are not just definitional choices; they determine which cases fall into which categories, which in turn determines the composition of the dataset, which in turn affects whatever statistical or comparative analysis follows. Two researchers studying "democracy" with different concept structures will literally be studying different sets of countries, even if they use the same word. The concept structure is not just a philosophical abstraction; it changes your results.

The negative pole

Goertz (2006) makes a point that we do not talk about enough: defining a concept also means defining what it is *not*. The negative pole of a concept (what the absence of it looks like) is as important as the positive pole. If your concept is "democracy," what is the non-democracy category? Is it autocracy? Anocracy? A residual category of everything that is not democratic? The content and coherence of the negative pole matters because in most research designs, you are comparing instances of the concept to non-instances. If the non-instance category is incoherent (if it lumps together cases that have nothing in common except not being democratic), then any comparison across the boundary is analytically weak. We need to ask ourselves: *what does my concept look like when it is absent?*

Connecting Goertz (2006) to Adcock and Collier (2001)

I think these two approaches are complementary rather than competing. Adcock and Collier (2001) tell you how to move from a background concept to a systematised concept to indicators to scores (and back again). Goertz (2006) tells you how to evaluate whether the systematised concept is internally well-structured, whether the dimensions are clearly specified, and whether the logical relationship between them is appropriate. Put differently, Adcock and Collier (2001) give us the *vertical* logic (moving down from abstraction to observation), and Goertz (2006) gives you the *horizontal* logic (how the components at each level relate to each other).

3. Collier and Levitsky (1997)

If Adcock and Collier (2001) give you the framework and Goertz (2006) gives you the toolkit, Collier and Levitsky (1997) show us what happens in practice when concept formation goes wrong and the creative but often problematic strategies academics have developed to cope with this. The article explores the proliferation of democratic subtypes that scholars have come up with in trying to apply it to an increasingly diverse set of political regimes. The result is a great example of conceptual analysis that is also a cautionary tale.

The problem: conceptual stretching

The article begins with a problem first identified by Sartori in the 1970s. As comparative politics expanded beyond Western Europe and North America, scholars began applying the

concept of democracy to a much wider range of political regimes. But the classical definition of democracy (i.e., competitive elections, civil liberties, rule of law, and accountability) did not fit many of the new cases well. Rather than saying that some regimes were simply not democracies, academics stretched the concept to accommodate them, often by dropping or weakening one or more of the defining attributes. The result was *conceptual stretching*: the concept of democracy became so broad that it lost analytical precision. If virtually every regime can be classified as some kind of democracy, the concept ceases to distinguish anything from anything else.

Sartori's ladder of abstraction

Collier and Levitsky (1997) adopt Sartori's (1970) ladder of abstraction. As you move up the ladder, concepts become more general (covering more cases) but thinner (with fewer defining attributes). As you move down, concepts become more specific (covering fewer cases) but thicker (with more defining attributes). There is, therefore, a trade-off: you cannot simultaneously maximise generality and precision. The classical concept of democracy sits at a particular rung on this ladder. Scholars who want to apply it more broadly can move up (adopting a thinner definition), and scholars who want more analytical precision can move down (adding more attributes). Neither move is inherently wrong, but each has consequences for what the concept can and cannot do in a research design.

Three strategies scholars use

This article catalogues three principal strategies that scholars have used to cope with the problem of applying democracy to a variety of regimes, each with distinct consequences.

The first strategy is *moving up the ladder* to a more general concept. Some scholars adopt a "minimalist" definition of democracy (for instance, defining democracy purely in terms of competitive elections). This increases the number of cases covered but strips the concept of attributes that many consider essential. The risk is that the resulting category groups together regimes that are fundamentally different in their politics.

The second strategy is *moving down the ladder* by creating classical subtypes. A classical subtype adds adjectives to the root concept (i.e., "parliamentary democracy," "federal democracy," "consociational democracy"). The resulting subtype is an instance of democracy with additional features. Classical subtypes increase precision at the cost of generality. They describe a narrower set of cases in greater detail.

The third strategy, and the one Collier and Levitsky (1997) focus on most, is the creation of *diminished subtypes*. A diminished subtype is missing one or more of the defining attributes of the root concept (i.e., "illiberal democracy," "delegative democracy," "electoral authoritarianism." Unlike classical subtypes, which *add* attributes, diminished subtypes *subtract* them. The case is still called a democracy, but it lacks something in the full concept. This is interesting because it shows scholars trying to have it both ways: preserving the democracy label while admitting that the country does not fully meet the definition. In Goertz's (2006) terms, diminished subtypes represent a shift from a necessary-and-sufficient structure (where all attributes must be present) to something closer to a family resemblance structure (where some can be absent).

The scale of the problem

One of this article's most interesting findings is the sheer number of subtypes of democracy. When there are so many definitions to choose from, scholars often talk past each other when

they talk about democracy. One researcher's "democracy" is another's "electoral authoritarianism." The accumulation of knowledge (in a Lakatosian sense) becomes difficult because findings generated under different definitions are not directly comparable. This is a real-world example of the problem Adcock and Collier (2001) highlight: disagreements at the background concept level spill through the entire chain of measurement.

The connection to your research (i.e., why should we care?)
This article is worth reading not because all of you are studying democracy, but because the dynamics Collier and Levitsky (1997) describe can be in play with almost any concept we care about in political science. Students working on "corruption" face the same ladder-of-abstraction trade-offs. Students working on "political violence" face the same proliferation of subtypes. Students working on "state capacity" face the same risk of conceptual stretching. This article gives us a way of diagnosing problems in our own work. When one of you say "I'm studying political trust" or "I'm studying populism," the follow-up question I always think of is: which definition of this concept are you using, and what are the consequences of that choice?

Connections across readings
Collier and Levitsky (1997) can be connected directly to the other two readings. Their discussion of diminished subtypes illustrates Goertz's (2006) point about concept structure: a diminished subtype is what you get when you shift from an AND structure to an OR structure without acknowledging that you have changed the concept. Their catalogue of subtypes connects to Adcock and Collier's (2001) point about background concepts diverging. The reason so many subtypes exist is that scholars are systematising the background concept of democracy in fundamentally different ways.

Discussion questions

Honours students

1. What are the one or two core *concepts* in your research design memo? Can you state, in a single sentence, what you mean by each one? If someone else here is using the same term differently, how would you resolve the difference?
2. Adcock and Collier (2001) distinguish between a *background* concept and a *systematised* concept. Take one of the key concepts from your research: what is the background concept (the broad set of meaning), and what is your systematised concept (the specific definition you are using)? What have you included and what have you excluded, and why?
3. Collier and Levitsky (1997) show that scholars often stretch a concept to cover cases it was not originally designed for. Can you think of an example from your own research interests where a concept has been stretched? What were the advantages or disadvantages of doing so?
4. Goertz (2006) argues that when a concept has multiple dimensions, you need to decide whether the dimensions relate by AND (all must be present) or OR (some can be absent). For one of your core concepts, which structure is more appropriate, and what would change about your research if you chose the other?

MA/PhD students

1. Adcock and Collier (2001) argue that measurement validity disputes often arise because scholars are working at different levels of the concept–measurement framework without realising it. Can you identify a dispute in your own research area where this diagnosis applies? At which level does the real disagreement sit, and what would resolving it require?
2. Collier and Levitsky (1997) identify diminished subtypes as a strategy for handling cases that do not fit the classical definition. In Goertz's (2006) framework, what structural move does this represent? Is it ever analytically legitimate, or does it always involve conceptual stretching?
3. Goertz (2006) emphasises the importance of the negative pole. For your own key concept, what is the negative pole? Is the set of non-instances coherent (i.e., a meaningful category in its own right), or is it a random assortment? What does this mean for your comparison or research design?
4. Sartori's (1970) ladder of abstraction implies a trade-off between generality and precision. Where on the ladder does your own project sit? What would change if you moved one rung up (making the concept more general) or one rung down (making it more specific)? Which direction would strengthen your design, and why?

Cross-reading questions (everybody)

1. Collier and Levitsky (1997) found over 550 subtypes of democracy. Using Adcock and Collier's (2001) framework, explain why this number of types occurred. At which level of the concept–measurement chain did the problem originate? Could it have been avoided, and if so, how?
2. Last week, Hyde (2007) operationalised "election fraud" as incumbent vote share at the polling-station level. Using this week's readings, evaluate that conceptual choice. What background concept of fraud is Hyde (2007) working with? What has she included and excluded? Would Goertz (2006) say her concept has an AND or an OR structure? What are the consequences for her findings?
3. Booth et al. (2016) argue that a research question becomes a research problem when you can articulate what is at stake. This week's readings suggest that a concept becomes a research tool only when you can articulate what it includes and excludes. How are these two arguments related? Can you have a well-formed research problem built on a poorly formed concept?

PART 3: GROUP WORK

Putting your concepts under pressure

This activity is designed to connect the readings directly to your research design memos before you submit at the end of the week. Please divide into pairs or small groups (~3) of either honours or HDR students. Each student identifies the one or two core concepts in their project and presents them to their partner(s). The partner(s) then ask the following questions:

1. Can you ***define the concept*** in one sentence without using the concept itself? (If you cannot, the concept is under-defined.)
2. Using Adcock and Collier's (2001) framework: is this a ***background*** concept or a ***systematised*** concept? Did you select a specific type from the broader set of meanings, or are you working with a vaguer version?

3. Using Goertz's (2006) framework: what are the ***dimensions*** of the concept? Are they related by AND or OR? Would you define the dimensions differently, and if so, what cases would be brought in or out?
4. Can you identify a clear ***example*** of the concept and a clear case that is not? What about a ***borderline*** case? How does your definition handle the borderline case?
5. Building off of Collier and Levitsky (1997), is the concept at risk of conceptual stretching? Are you applying a concept developed in one context to a very different context? If so, what attributes might not travel?
6. How does the definition of the concept affect which cases you will study? If you changed the definition, would your cases (and findings) change?

PART 4: YOUR DESIGN MEMO

The design memo is due Friday 11:59pm this week. This is the last class before submission.

Remember what the memo asks for:

- A clear research question
- An explanation of its significance
- The theoretical framework and associated hypotheses
- A description of the proposed methodology including data sources (if possible), analytical strategies, competing designs, and scope conditions
- Identification of key risks
- Consideration of alternative methodological approaches

A note on ambiguity
In my experience, students often feel they need to project confidence in their research design and seem to be leery of admitting uncertainty about a concept or definition. I want to be clear that the opposite is true. A research design that says "I am using X's definition of corruption, which focuses on the abuse of public office for private gain, because my data sources (court records and audit reports) are best suited to capturing this type while acknowledging that this excludes patronage and clientelism, which other scholars would include" is ***stronger*** than a memo that simply says "I am studying corruption." Clarity and precision about what you are doing and what you are not doing is a sign of conceptual maturity, not weakness. Adcock and Collier's (2001) framework gives us permission to make these choices explicitly and defend them.

Looking ahead
Next week turns from concept formation to operationalisation and measurement: how the concepts students have defined this week get translated into variables, indicators, and data. The readings will address measurement validity and reliability in more depth, building directly on the Adcock and Collier (2001) framework introduced this week.