21

# Causal Inference

## *The Critical Role of Concepts*

Thad Dunning

### INTRODUCTION

The study of political methodology places substantial emphasis on causal inference. This seems appropriate. After all, many or most social science theories focus on relations of cause and effect. The problem of drawing valid inferences about causal relations from empirical research is thus central to social science.

Unfortunately, however, the study of conceptualization and measurement appears to play a more minimal role in contemporary methodological research; especially, it occupies a relatively minor place in many graduate methods programs.[1] This is unfortunate for two reasons. First, strong concepts and measures are the foundation of many forms of empirical research not necessarily linked to causal inference – including centrally description. Although inferring causation requires description, what is sometimes called "descriptive inference" plays an important role beyond causal inference. Second – and most relevant to this chapter – successful concept formation is itself a *sine qua non* of the successful study of causation. The neglect of concept formation in graduate curricula can thus undermine the effort to improve the quality, validity, and meaningfulness of causal inference.

Consider experiments, a key tool for causal inference. The design of an experiment immediately raises important conceptual questions. What is the "treatment" a case of – that is, what concept does it instantiate? To what class of similar treatments is it comparable? In what other empirical settings might we expect its effects to be found? As I argue in this chapter, these

[1] This is not to gainsay excellent contemporary work on concepts and measurement, some of which is discussed in this chapter and the volume of which it is part. See also Gary Goertz (2020).



Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 09:52:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.032

questions of both internal and external validity are matters not only of theory and empirics but very centrally of concepts.

The tools of concept formation are therefore essential. In this chapter, I discuss how to apply Giovanni Sartori's ladder of abstraction to problems of causal inference, with particular emphasis on experimental design. As I suggest, Sartori's ideas and related innovations in other foundational work on concepts, especially Collier and Levitsky (1997),[2] are highly relevant for cutting-edge problems of causal inference.

## SARTORI'S LADDER OF ABSTRACTION

Giovanni Sartori's (1970) path-breaking article "Concept Misformation in Comparative Politics" made two crucial contributions.

First, it underscored the critical role of conceptual classification. Sartori wrote in a context in which some researchers discredited taxonomies – in particular, dichotomies – as primitive and pre-quantitative. Measurement, for those researchers, began with graded or especially interval-level scales.[3] But Sartori rightly emphasized instead the key role of classification: Before deciding "how much" of a thing there is, scholars must assess whether the thing is an instantiation of a particular concept or not – that is, in what conceptual container it belongs. This cannot be done without the work of elaborating the properties of the concept. As Sartori (1970: 1038) put it, "concept formation stands prior to quantification."

Second, Sartori ingeniously showed how the "traveling problem" that bedevils much empirical social science, and that is perhaps most accentuated in comparative politics, is fundamentally a conceptual challenge. He proposed his famous ladder of abstraction[4] as a way to address it (see Figure 1 in Collier and Levitsky 1997). The relationship between the *extension* (or denotation) of a concept and its *intension* (or connotation) is key. Extension refers to the specific units to which a concept applies: These particular countries are democracies; those individuals are bureaucrats. Intension instead specifies the properties necessary for membership in a conceptual category: for instance, democracies have (1) free and fair elections, (2) a certain level of popular participation, and (3) liberal protections for minority rights.

Climbing the ladder of abstraction involves sacrificing intension for greater extension – for instance, by replacing "democracy" with "regime," which applies to a greater range of countries. It thus achieves greater generality, without losing precision: at a higher level of abstraction, it is still conceptually clear which instantiations should count as members of a class and which should not. Scaling the ladder of abstraction is thus a way of achieving greater

[2] See Chapter 2, this volume.
[3] For a discussion of these critiques, see Chapter 7, this volume.
[4] Collier and Levitsky (1997) call this a ladder of generality – intending the same meaning.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 09:52:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.032

generality while avoiding conceptual stretching, though it can come at the cost of losing conceptual differentiation (Collier and Levitsky 1997). Thus, for many purposes, *regime* is simply too general a term. Conversely, broadening extension without reducing intension can produce conceptual stretching. For example, it may involve labeling countries as democracies, even though their attributes do not make them members of the class.

To be sure, as Collier and Mahon (1993) emphasized,[5] Sartori's idea of conceptual stretching is anchored in "classical" categories with clear boundaries. The ideas of family resemblances and "radial" conceptual structures can relax this constraint. Specifically, one way to increase conceptual differentiation without stretching is thus with diminished subtypes, for example, "democracy with adjectives" (Collier and Levitsky 1997). An illustration would be "illiberal democracies," which lack protection for minorities and thus are not full instantiations of the root concept of "democracy." Crucially, Sartori's approach and the use of diminished subtypes may be viewed as complementary procedures; in other words, they can be used together.

The ladder of abstraction provides guidance for the formulation of concepts that travel. Consider "staff," "administration," and "civil service" as alternative labels for the set of agents who implement the directives of governments or rulers. Sartori (1967: 1042) quotes Smelser (1967: 103), who argues that "staff is more satisfactory than administration … and administration is more satisfactory than civil service … the concept of civil service is literally useless in connection with societies without a formal state or governmental apparatus … the concept of administration is somewhat superior … but even this term is quite culture-bound." Consequently, Sartori suggests that – again quoting Smelser – the more useful, broader term is "Weber's concept of staff … since it can encompass without embarrassment various political arrangements."

Thus, a key component of generalization involves the choice of the level of abstraction at which units are comparable in this sense – that is, at which the root concept is sufficiently high on the ladder of abstraction to generate conceptual homogeneity while still maintaining differentiation.

## TWO CONCEPTUAL CHALLENGES FOR EXPERIMENTAL RESEARCH

I now turn to the relevance of Sartori's contributions for experimental research. I focus on their usefulness in addressing just two of many important challenges.

### Challenge 1: What Is the Treatment?

One core challenge in experimental research is defining the conceptual "container" into which a treatment should be placed – that is, the concept of which

[5] See Chapter 1, this volume.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 09:52:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.032

it is an instance. With experiments, the question "What is the treatment a case of?" may thus often be asked.[6] This challenge clearly pertains to observational studies as well. Yet it may be particularly noticeable to experimental researchers (and their audience), given their greater control over the design of an intervention.

When experiments seem unsatisfying or uninformative, I believe, it is often because the proper placement of the experimental treatment on a relevant ladder of abstraction does not match the level of generality of a core concept in the theory. This could happen for at least three reasons. First, the experimental treatment is lower on a ladder of abstraction than the root theoretical concept. Second, it may also contain only some of the attributes that give the root concept in a theory its connotation, and the missing attributes appear theoretically important. In this case, one might think of a treatment as if it were a diminished subtype – lacking some properties of a middle- or higher-level concept. Finally, a concept is applied improperly to a given experimental treatment in a manner akin to conceptual stretching.

The extent to which this problem arises varies across experiments. As one example, in the outstanding study of Hainmueller et al. (2018), researchers designed two randomized experiments to encourage eligible immigrants in New York to attain US citizenship. In one experiment, certain low-income, lawful permanent residents registered for a public/private naturalization program. Yet they were ineligible for a federal program waiving the $680 application fee.[7] Some of these applicants were offered – at random – remission of the fee. The study found that the subsidies to eliminate financial costs raised the naturalization application rate by 42 percent.

By contrast, in a second experiment, even lower-income applicants, who were already eligible for the federal fee remission, received behavioral nudges designed to help them overcome nonfinancial hurdles in the naturalization process.[8] Nudges are an important topic in the literature, and they have received substantial attention from behavioral economists. These randomly assigned nudges – which comprised reminders, assistance, and encouragement – were similar to those used by service providers working with immigrant populations. As it turned out, nudges had no discernible effect on naturalization.

Especially from a policy perspective, the interventions in these two experiments correspond nicely to concrete instantiations of concepts such

[6] Here I distinguish this question from the common related problem of deciphering the "active" element of a treatment – though that is also an important problem that is partly conceptual.

[7] These applicants were lawful permanent residents (LPRs) who have household incomes between 150 percent and 300 percent of the federal poverty guidelines and who did not receive means-tested government benefits.

[8] The applicants in this second group were very low-income LPRs who had a household income below 150 percent of the federal poverty guidelines or received means-tested benefits.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 09:52:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.032

as "cost" for individual applicants. The nudges may also correspond to improvements in (perhaps a somewhat narrow conception of) legal receptiveness on the part of a host population.

At the same time, targeted reminders and encouragement may not correspond to a broader "context of reception" that, according to other scholars of immigration, may more deeply shape the attractiveness of citizenship to legal migrants (Portes and Rumbaut 2001; Fox and Bloemraad 2015).[9] Indeed, this broader context of reception may be invariant to whether or not respondents receive nudges from service providers – which could help to account for the null effects of nudges. This observation in no way gainsays the value of testing the impact of nudges per se. However, it does raise the question of the experiment's connection to core theoretical concepts in the study of barriers to immigrant naturalization. The point is especially relevant if such a broader background concept is indeed the theoretical quantity of greatest interest.

In this and many other studies, it is therefore useful to connect the intervention to a systematized, root, or background concept. To interpret an experiment's results, a core task is one of classification – is this treatment an instance of the concept? This conceptual task can also be viewed from the prism of internal validity – of whether "experimental treatments [did in fact] make a difference in this specific experimental instance" (Campbell and Stanley 1963: 5). To interpret any difference, one must conceptualize the "difference maker" or causal agent at work.

The tools of concept formation are helpful for this critical task. The ladder of abstraction is conceptually clarifying and connects the empirics to relevant theory. For example, one could reformulate a root concept to cast it at a lower level of abstraction more appropriate to an experiment. Relatedly, scholars may define a new version of a systematized concept – a "specific formulation of a concept adopted by a particular researcher" (Adcock and Collier 2001: 530),[10] thereby ignoring some attributes associated with a broader background concept. Better conceptualization can produce a more satisfying connection between theory and empirics. Of course, it can also render interpretation more circumspect, by potentially laying bare the limited theoretical meaningfulness of a given experiment.

Careful conceptualization of the treatment is thus a critical aspect of experiments and of observational research oriented toward causal inference.

## Challenge 2: Cumulative Learning

Experiments are often prized for the purchase they provide for causal attribution in a single study – their internal validity.

[9] To be clear, Hainmueller et al. (2018) rightly do not interpret the behavioral nudges as instantiations of a context of reception, but they do discuss that concept as motivation for the research.

[10] See Chapter 16, this volume.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 09:52:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.032

Recently, however, experimental researchers have begun to address more centrally and systematically the challenge of cumulation of knowledge. Results from different experiments may fail to aggregate into a meaningful set of cumulative findings for many reasons. First, there are few studies on a particular topic. Once the "flag" has been planted by a set of researchers, incentives for replication are weak. Second, studies differ substantially in their designs, complicating the pooling or comparison of results. Finally, many studies in which null effects are found are not reported, making the conclusions drawn from a set of published studies potentially problematic. Generalizability (or Campbell's "external validity") is one core aspect of this problem, but the challenge for the cumulation of knowledge is broader.

The Metaketa Initiative spearheaded by the Evidence in Governance and Politics group (Dunning et al. 2019) is one recent effort that seeks to address this challenge of cumulative learning. To address problems of study scarcity, study heterogeneity, and selective reporting, researchers have collaborated on the design of studies across divergent contexts, preregistering a meta-analysis of results before the studies are conducted. In our inaugural collaborative project, my coauthors and I sought to assess the impact of providing voters with positive or negative information about incumbents' political performance (Dunning et al. 2019). Each of the studies included in the preplanned meta-analysis were randomized controlled experiments, with informational interventions randomized across respondents in each case.

Yet the pooled analysis of an intervention "common" to each study raises again the critical question noted earlier of study heterogeneity. How can heterogeneity conceivably be reduced across a set of contexts as disparate as Benin, Brazil, Burkina Faso, India, Mexico, and Uganda? How, in particular, can one homogenize the "main" treatment arm in the set of studies in which the details of incumbency and the positive and negative information differed nontrivially across these contexts? This is an important challenge for all meta-analysis.

Sartori's conceptual tools provide an immediately helpful resource for addressing such questions. Indeed, Sartori focused himself on the challenge of cumulation.[11] The challenge of comparability suggests Sartori's "traveling problem" and thus brings into focus the relevance of the ladder of abstraction. The question of the differences of the treatments across the diverse study sites can therefore be productively recast as a question of abstraction – that is, the level on Sartori's ladder at which the treatments

[11] For instance, Sartori (1970: 1043) warns against "research … not guided by [a] universal contract but by intuitive concrete conceptualizations … The unfortunate consequences are not only that the research lacks theoretical backing (for want of medium level categories, and especially of a taxonomic framework) but that the vagueness of the theory has no fit for the specificity of the findings. We are thus left with a body of literature that gives the frustrating feeling of dismantling theoretically whatever it discovers empirically."

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 09:52:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.032

can be productively "homogenized" as a conceptual matter and thus their effects can be meaningfully compared or integrated.

In our case, at the lowest levels of abstraction, it was clear that any two interventions using different kinds of information and taking place in contexts as distinct as six different countries in Latin America, Africa, and South Asia must differ on a very large number of dimensions (Dunning et al. 2019). As the attributes needed to define a "common" intervention multiply, the number of treatments to which a concept can apply diminishes – illustrating Sartori's trade-off between intension and extension. Yet by focusing on just a few core attributes of commonality and thereby climbing the ladder of abstraction, the generality and extension of the concept may increase. A central focus in our common study of informational interventions was the distinction between "good" and "bad" news about political candidates.[12] This concept is at a middle level of abstraction and can be meaningfully defined across disparate contexts, even if the particularities of the interventions differ substantially. If theoretical expectations are defined at this level of abstraction, then empirical aggregation of results can also be feasible and meaningful.

Whether this effort was successful in this specific instance is for others to judge. The key point here is that the conceptual considerations suggested by Sartori's ladder of abstraction proved essential.

## CONCLUSION

Concept formation is critical for successful causal inference, as it is for other social-scientific goals. I have focused here on experiments, but the observations can readily be extended to cognate designs, such as natural experiments, or to mainstream observational studies (Brady and Collier 2010).

We would therefore do well to heed Sartori's admonition that concept formation stands prior to quantification – but we can also replace "quantification" with "experimentation," or indeed more broadly with "causal inference."

## REFERENCES

Adcock, Robert, and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95(3): 529–46. (See Chapter 16, this volume.)

Brady, Henry E., and David Collier. 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman & Littlefield.

Campbell, Donald T., and Julian C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin.

[12] This we operationalize by studying the relationship between voters' prior beliefs and the information that is provided to them about candidate performance.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 09:52:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.032

Collier, David, and Steven Levitsky. 1997. "Democracy with Adjectives: Conceptual Innovation in Comparative Research." *World Politics* 49(3): 430–51. (See Chapter 2, this volume.)

Collier, David, and James E. Mahon. 1993. "Conceptual 'Stretching' Revisited: Adapting Categories in Comparative Analysis." *American Political Science Review* 87(4): 845–55. (See Chapter 1, this volume.)

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan D. Hyde, Craig McIntosh, and Gareth Nellis (eds.). 2019. *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*. New York: Cambridge University Press.

Fox, Cybelle, and Irene Bloemraad. 2015. "Beyond 'White by Law': Explaining the Gulf in Citizenship Acquisition Between Mexican and European Immigrants, 1930." *Social Forces* 94(1): 181–207.

Goertz, Gary. 2020. *Social Science Concepts and Measurement*. Princeton, NJ: Princeton University Press.

Hainmueller, Jens, Duncan Lawrence, Justin Gest, and David Laitin. 2018. "A Randomized Controlled Design Reveals Barriers to Citizenship for Low-Income Immigrants." *Proceedings of the National Academy of Sciences* 115(5): 939–44.

Portes, Alejandro, and Rubén G. Rumbaut. 2001. *Legacies: The Story of the Immigrant Second Generation*. Berkeley and Los Angeles: University of California Press.

Sartori, Giovanni. 1970. "Concept Misformation in Comparative Politics." *American Political Science Review* 64(4): 1033–53.

Smelser, Neil J. 1967. "Notes on the methodology of comparative analysis of economic activity." *Social Science Information* 6(2–3): 7–21.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 09:52:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.032