

Capturing causation in political science: the perspective of research design

Luigi Curini
Alessia Damonte*

Dept. of Social and Political Sciences,
Università degli Studi di Milano, Milano, Italy

Green version of

Curini L, Damonte A (2021). Capturing causation in political science: the perspective of research design. *Italian Political Science Review/Rivista Italiana di Scienza Politica* 1–7. <https://doi.org/10.1017/ipo.2021.28>

Abstract

In the last decades, ‘research design’ has become a strategic topic across political science. An emerging discourse relies on it to encompass paradigmatic oppositions and cultivate a pluralist approach to causation. As an introduction to the special issue on the topic, we offer an outline of the roles that the discipline recognizes to design in its relation to models and contend that, in a time of fascination for predictors, political science pluralism allows for balancing interpretability and validity of findings at once.

Key words

Causal inference; causal model; political science; research design

Introduction

Once the hallmark of the experimental approach (e.g., Cox and Reid, 2000), ‘design’ has come to refer to any research planning deployed to ensure sound results (Creswell and Creswell, 2017). In political science, the term gained traction after King *et al.* (1994) proposed a template of reference for comparative politics and international relations with the explicit motivation to fill a gap. At the time, they found that young scholars were ‘highly sophisticated qualitative and quantitative data collectors, interviewers, soakers and poakers, theorists, philosophers, formal modelers, and advanced statistical analysts,’ yet often ‘had trouble defining a research question and designing

the empirical research to answer it.' On top of that, students 'proposed impossible fieldwork to answer unanswerable questions' (King *et al.*, 2004: 112). Their proposal conveyed the image of a discipline to which causation was the driving concern, experiments provided the standard of reference, and the statistical treatment of large sets of data offered the viable approximation worth pursuing.

This strategy for methodological consolidation is far from uncommon in the social sciences and seldom goes unchallenged. A few decades before, the experimental design had been set as the benchmark for program evaluation (Campbell and Stanley, 1963; Cook and Campbell, 1979). In response, the unsympathetic methodological quarters entrenched beyond paradigmatic lines and causal analysis became a disputable practice (e.g., Guba and Lincoln, 1989). Political science tells a different story. Despite heterogeneity (e.g., Mahoney and Goertz, 2006; Della Porta and Keating, 2008), the fields of comparative politics and international relations aimed for common ground. Later influential contributions agreed – with some reasonable *caveat* – that credible causal inference provides a common motivation in political science and bridges research across techniques and traditions (e.g., Gerring, 2001; Collier *et al.*, 2004: 57 ff; Ragin, 2008; Rohlfing, 2012; Blatter and Haverland, 2012; Bennett, 2015; Fairfield and Charman, 2017; also Levy, 2007). A major consequence of such focused pluralism lies in the emerging practices of mixing (e.g., Brannen, 2005; Onwuegbuzie *et al.*, 2009), nesting (e.g., Lieberman, 2005; Rohlfing, 2008), and triangulating (e.g., Hammersley, 2008) techniques from different traditions into the same study. Meanwhile, the call for replicability and transparency has enhanced the relevance of design choices to establish the credibility of research findings (King, 1995; Blair *et al.*, 2019).

Since then, the methodological discourse has been compelled to provide meaningful guidance on conducting sound causal research and preserve the dialogue across studies. This special issue follows from the seminar that, with this aim, the Standing Group in Research Method in Political Science – MetRiSP held in 2019 at the University of Catania under the auspices of the Italian Society of Political Science – SISP. The seminar shared the tenet that, on pain of demoting research to homiletics, credible explanations of political and social phenomena entail the empirical testing of causal claims (i.e., Ermakoff, 2019) – although, in turn, testing produces a new crop of problems. Some of them have a technical nature and depend on the toolbox of choice: frequentists are concerned by statistical power and significance; Bayesians, by decisions about priors; set algebraists, by calibration settings. Other problems are common conceptual issues – such as deciding how to define the phenomena of interest or which manifestations capture them without stretching the definition. In between, the problem lies in establishing the causal role of a factor. Here, design enters as the default warrant of credible

ascription.

Research design is the rationale for connecting evidence to a causal model to establish its tenability – the major difference lying in model specification. Designs, therefore, warrant the pluralist *desiderata* of defining traditions in causal research while avoiding estrangement, to ease the cumulation of substantive causal knowledge.

How political science understands research design

Across the discipline, the methodological discourse offers four major definitions of design, as follows.

Design as a stage in the research cycle

Schmitter (2008: 294) makes sense of research design as the solution to an implementation problem. He portrays the research process as a cycle of strategic decisions, clustered into 11 themes and four stages. The stage of 'discovery' entails decisions on the topic, its conceptualization, and the generation of a hypothesis as an 'if-then' conditional. The hypothesis initiates the second stage of 'explication,' where decisions are made on selecting cases and operationalizing variables – and the proposal is written. The operationalization opens the third stage of 'accuracy,' in which measures are decided and tested as needed. To Schmitter, this also is the time for serendipitous reconsiderations, as measurement may retrofit concepts. The last stage, of 'proof,' provides the technique of choice with the empirical basis for establishing the tenability of the 'if-then' claim. The issue of design arises during explication as the question of which observations provide a proper source of evidence (*ivi*, 276). The answer is laid on the degree of control that the researcher can exert on the variation of key factors. Given fully manipulable variables, the selection can pursue a pure experimental design. Without manipulability, but given the possibility of simulating it, the selection should embark on quasi-experiments. Absent that possibility, those researchers affording reliable cross-unit observations can opt for comparisons and case studies. Given no empirical basis, the researcher can still resort to thought experiments – either rhetorical, or counterfactual as in the analytical narrative, the spatial theory of voting, or agent-based modeling approaches, just to mention three possibilities (Austen-Smith and Banks, 2005; McCarty and Meirowitz, 2007; Martelli, 2009; Laver and Sergenti, 2011).

Design as a defining feature of empirical studies

Toshkov (2016) engages in arranging many renowned distinctions in the methodological literature within a single taxonomy of research. He identifies 'descriptive,' 'predictive,' and 'explanatory' as the three *genera* within the family of empirical research that, with the theoretical family, compose the 'positive' research order as distinct from the 'normative' one. Then, he introduces further distinctions between species of positive studies: experimental and observational, statistical large-N and comparative small-N, cross-case and within-case.

Like in Schmitter's portrayal, positive studies are the branch in the taxonomy where the actual issue of designing arises (*ivi*, 168). In a partial departure from the previous proposal, however, here the relevant diversity depends on both the case selection and the manipulability of the quantity of interest, as independent design choices set the two. Case selection is understood as the possibility of random assignment of the units of observations to a state of the variable of interest, independent of the possibility of experimental control over that state.

Thus, the 'gold standard' of randomized control trials results when both random assignment and experimental control are employed in the same design. Experimental control alone allows for quasi-experiments, while random assignment alone makes natural experiments possible. When no control is applied, the study becomes observational.

Design as an approach to causal inquiry

The third perspective comes from Gschwend and Schimmelfennig (2007), who connect design to types of causal inference.

Their typology makes designs follow from the number of observations, opposing large-N vs. small-N types of studies. Variation in numerosity indicates alternative renderings of causation – deterministic or probabilistic – and of challenging observations – interesting 'deviant cases' or noisy 'outliers' (*ivi*, 13). The second dimension contrasts 'factor-centric' and 'outcome-centric' designs. The authors borrow the distinction from George and Bennett (2004) and the case study tradition but assume its general applicability (indeed: Holland, 1986). The distinction points to two kinds of questions to which research can respond – namely, whether a single factor of interest can be proven relevant to an outcome, or which bundle of factors account for the outcome of interest. They consider that picking one question has practical consequences on design. Factor-centered research pursues the insulation of the relationship of interest to magnify its signal against others that may interfere with it. Instead, outcome-centered research aims to identify the complex of the relevant relationships to some actual outcome and widens to selected interplays. Thus, they reason, the

first strategy establishes the predicting capacity of a factor; the latter, its explanatory value.

The combination of the two dimensions – numerosity and research question – yields four types of designs (*ivi*, 14). Pure or natural experiments, and studies affording statistical control, illustrate the factor-centric large-N type, while cross-case comparisons and quasi-experiments are factor-centric small-N. On the side of the outcome-centric types, forecasting and Qualitative Comparative Analysis are examples of the large-N sort, while case studies identify the small-N ones.

Design as a diagnostic device

The proposal by Blair *et al.* (2019) responds to the problem of how to ‘declare’ one’s research design so that the information conveys the merits of the researcher’s choice against alternative specifications.

The framework serves the purpose of easing peer scrutiny and, retrospectively, guiding design decisions. Their examples show that the scheme applies to descriptive designs (e.g., surveys, Bayesian inference), causal designs (Process Tracing, Qualitative Comparative Analysis, Nested Analysis with qualitative confirmation, Matching on Observables, Regression Discontinuity, and experimental), and discovery-oriented designs (latent allocation). However, the authors acknowledge that a ‘design declaration’ may become complete at different stages of the research process depending on the motivation of the study. They reason that early and complete declarations follow from testing hypotheses about single factors on secondary data. When inference serves discovery, the details of a design may become fully clear *ex-post* only; thus, they may not allow alternative specifications unless committing to a hypothesis.

Following Geddes (1990), they ponder that a complete design consists of four elements: a causal model of the world (\mathcal{M}), an inquiry (\mathcal{J}), a data strategy (\mathcal{D}), and an answer strategy (\mathcal{A}). \mathcal{M} includes the factors considered for the analysis and the assumptions about the shape and direction of the relationships they entertain. In short, \mathcal{M} is the data generation model to capture a pattern in the real world. \mathcal{J} is that which the researcher wants to learn about \mathcal{M} : either the conditional values of a special factor Y , or the values that the factor would take under intervention. \mathcal{D} refers to the strategies employed to construe evidence about \mathcal{M} that answers \mathcal{J} and include data collection, sampling, assignment, casing, and the mapping of latent variables onto observable ones. The answer \mathcal{A} is declared through the techniques deployed to turn data from \mathcal{D} into evidence responding to \mathcal{J} .

These elements, and their content, convey a broader understanding of the research design as the set of decisions made along the whole research

process: \mathcal{M} captures theory formulation, \mathcal{J} corresponds to the research question, \mathcal{D} indicates the criteria for case selection and coding, while \mathcal{A} is for the actual strategy for inference (ivi, 842). Although analytically discrete, the authors reason that these elements are connected by a relationship of dependence: the data strategy \mathcal{D} and the answering strategy \mathcal{A} follow from the interplay between a model \mathcal{M} and an inquiry \mathcal{J} (ivi, 852). The priority of \mathcal{M} and \mathcal{J} , the authors reason, holds even in discovery-oriented studies. Despite the fact that it may proceed from a less structured starting point, no empirical strategy can venture into the field without some sense-making device.

The hallmark of the pluralist approach

Although far from overlapping, the four definitions just discussed agree on the tenet that causal claims can be established credibly despite the fact that our ascription can be biased. Research design is for keeping our biases under control while gathering the evidence that decides the claims' tenability. The legitimate causal claims, importantly, may narrow on single causal factors or widen to causal structures.

The merit of the latter point becomes especially clear against the ongoing debate on the limits of the methodological consolidation project. The point of contention concerns the core belief that manipulation and random assignment provide some superior causal evidence as they warrant the independence of conclusions on theoretical assumptions about causal structures (e.g., Imbens, 2020). The limit consists of reducing the data generation model to some minimal stimulus-response mechanism, and causal analysis to an estimate of the difference in the aggregate response of the units exposed to the stimulus and their unexposed statistical twins. This strategy may provide a first answer to whether the stimulus works; however, it dismisses the related questions of how it happens and under which conditions it succeeds or fails as matters of disturbance of an information signal. These questions are far from ancillary. Even when understood as the simplest elicited response, success or failure always is a local matter of the stimulus meeting the right conditions or no obstructions (e.g., Pemberton and Cartwright, 2014, Kaaber, 2020). At the same time, the ascription of the width of the response to the stimulus remains credible unless confounders are ruled out. When such relevant conditions and confounders are not explicitly considered, they may bias the estimate of the stimulus's net effect in unknown directions (e.g., Pearl, 2009). Moreover, the four definitions discussed above show that the pluralist stance can place political science beyond the debate. They implicitly acknowledge that causal analysis always requires models as the selection of meaningful connections in a tangled world. Besides, relevant background conditions and

structural assumptions enter either as controls of the estimate of the single relationship of interest or as compound factors before an outcome (e.g., Franzese, 2007, Ragin, 2008, Bennett, 2015). Designs warrant the tenability of any causal model. An 'if-then' relationship hardly holds unless some 'if-else' counterfactual evidence is available that, had the factor been different under the circumstances, the response would have changed, too. Design tackles the problem of finding or construing the proper 'if-then-else' empirics to dispel the doubt that the relationship only lives in our imagination.

Along this line, in this issue, Valentim *et al.* (2021) offers a blueprint of the quasi-experimental design based on the discontinuity in the value of a factor. The strategy assumes that the units close to the threshold approximate experimental twins and allow estimating the average treatment effect of the factor, albeit with local validity. Valentim *et al.* (2021) shows how theory is still required to establish the plausible functional form of the effect and generalize the findings.

In a similar vein, Costalli and Negri (2021) introduce matching techniques to identify those units in the field that explicitly make the counterfactual claim observable. They discuss how these techniques find the twins in the units that display similar selected background features and how such a similarity can be estimated with different gauges. In a partial departure from estimates that rely on mapping the manifold background features onto a unidimensional propensity score, they illustrate the so-called coarse matching in which the relevant background features are turned into meaningful classes. Here again, theory is implicitly required to establish which background features are relevant and how classes are meaningful.

Di Salvatore and Ruggeri (2021) make the role of theory and models explicit in addressing a special source of threat to the credible estimate of the net effect, usually discounted through the error term. Against the standard assumption that the response follows from the direct exposition of a unit to the causal factor of interest, they contend that, in real settings, causation can also proceed indirectly from exposed units to their neighbors through 'transfer' mechanisms such as spillover or mimesis. Hence, they show how these spatial effects can be modeled and gauged to improve causal estimates.

Martini and Olmastroni (2021) go further in the direction of model specification as theory testing. They discuss the application of the experimental rationale to surveys, to test theories about people's preferences. They show how, in factorial and conjoint designs, an eliciting factor can be modeled as a special configuration of key features expected to interact with special traits of the respondents, which they recognize as moderating factors. The random assignment of actual respondents to the eliciting factor's components allows testing the hypothesis while freeing responses from social desirability bias.

Damonte (2021) again applies the configurational understanding of the causal

factor in a quasi-experimental design, although under the assumption that the causal factor is the whole compound of right conditions. The causal analysis is geared toward pinpointing the bundle of theoretically meaningful factors that support an 'if-then-else' claim about an outcome beyond the rationale of the net effect. She shows set-theoretical techniques and logical pruning operations can identify the relevant compounds beneath each state of the outcome, and that the compounds retrieved under alternative counterfactual assumptions can constitute mediated causal structures.

Ruffa and Evangelista (2021) discuss the process-tracing approach to the data-generation model as a theory-driven definition of the causal situation and a hypothesis about the special unfolding of a chain of events before an outcome. This chain turns hypothesis testing into the retrieval of the marks or fingerprints that the chain of events would have left in the cases, were the hypothesis true. They recognize the challenge of the technique in identifying meaningful marks, establishing the evidential value that each can bear, and converting that value into a weight for the Bayesian update of our beliefs in favor or against the tenability of the hypothesis in the case against its alternatives.

Together, the articles attest that the 'if-then-else' understanding of causation embodied in the quasi-experimental rationale affords plural renderings, each suitable to respond to specific questions. The contributions, moreover, suggest dismissing the debate between the primacy of model vs. design as the one cannot really dispense from the other. The ultimate rift in research may run less between techniques, languages, or questions than between quarters – those rejecting causation as a legitimate and fruitful object of the discipline and those embracing the challenge of better causal knowledge.

References

- Austen-Smith D and Banks JS (2005) *Positive Political Theory II: Strategy and Structure*. Michigan: Michigan University Press.
- Bennett A (2015). Disciplining our conjectures: systematizing process tracing with Bayesian analysis. In Bennett A and Checkel JT (eds), *Process Tracing: From Metaphor to Analytic Tool*. Cambridge, UK: Cambridge University Press, pp. 276–298.
- Blair G, Cooper J, Coppock A and Humphreys M (2019) Declaring and diagnosing research designs. *American Political Science Review* 113, 838–859. doi:10.1017/S0003055419000194
- Blatter J and Haverland M (2012) *Designing Case Studies: Explanatory Approaches in Small-N Research*. London, UK: Palgrave Macmillan.
- Brannen J (2005) Mixing methods: the entry of qualitative and quantitative approaches into the research process. *International Journal of Social Research Methodology* 8, 173–184. doi:10.1080/13645570500154642
- Campbell D and Stanley J (1963) *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL: Rand-McNally.
- Collier D, Seawright J and Munck GL (2004) The quest for standards: King, Keohane, and

- Verba's 'Designing Social Inquiry'. In Brady HE and Collier D (eds), *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Plymouth, UK: Rowman and Littlefield, pp. 33–63.
- Cook TD and Campbell DT (1979) *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin Company.
- Costalli S and Negri F (2021) Looking for twins: how to build better counterfactuals with matching. *Italian Political Science Review/Rivista Italiana di Scienza Politica*, 1–16. <https://doi.org/10.1017/ipo.2021.1>
- Cox DR and Reid N (2000) *The Theory of the Design of Experiments*. Boca Raton, FL: Chapman & Hall/CRC Press.
- Creswell JW and Creswell JD (2017) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. London, UK: Sage.
- Damonte A (2021) Modeling configurational explanations. *Italian Political Science Review/Rivista Italiana di Scienza Politica*, 1–16. <https://doi.org/10.1017/ipo.2021.2>
- Della Porta D and Keating M (2008) How many approaches in the social sciences? An epistemological introduction. In Della Porta D and Keating M (eds), *Approaches and Methodologies in the Social Sciences A Pluralist Perspective*. Cambridge, UK: Cambridge University Press, pp. 19–39.
- Di Salvatore J and Ruggeri A (2021) Spatial analysis for political scientists. *Italian Political Science Review/Rivista Italiana di Scienza Politica*, 1–17. <https://doi.org/10.1017/ipo.2021.7>
- Ermakoff I (2019) Causality and history: modes of causal investigation in historical social sciences. *Annual Review of Sociology* 30, 581–606. doi: 10.1146/annurev-soc-073117-041140
- Fairfield T and Charman A (2017) Explicit Bayesian analysis for process tracing. *Political Analysis* 25, 363–380. doi:10.1017/pan.2017.14
- Franzese R Jr (2007) Multicausality, context-conditionality, and endogeneity. In Boix C and Stokes SC (eds), *The Oxford Handbook of Comparative Politics*. Oxford, UK: Oxford University Press, pp. 27–72.
- Geddes B (1990) How the cases you choose affect the answers you get: selection bias in comparative politics. *Political Analysis* 2, 131–150. doi:10.1093/pan/2.1.131
- George AL and Bennett A (2004) *Case Studies and Theory Development in the Social Sciences*. Boston, MA: The MIT Press.
- Gerring G (2001) *Social Science Methodology. A Criterial Framework*. Cambridge, UK: Cambridge University Press.
- Gschwend T and Schimmelfennig F (2007) Introduction: designing research in political science – a dialogue between theory and data. In Gschwend T and Schimmelfennig F (eds), *Research Design in Political Science: How to Practice What They Preach*. Basingstoke: Palgrave MacMillan, pp. 1–18.
- Guba EG and Lincoln YS (1989) *Fourth Generation Evaluation*. London: Sage.
- Hammersley M (2008) Troubles with triangulation. In Bergman M (ed.), *Advances in Mixed Methods Research: Theories and Applications*. Thousand Oaks, CA: Sage, pp. 22–36.
- Holland PW (1986) Statistics and causal inference. *Journal of the American Statistical Association* 81, 945–960. doi:10.2307/2289064.
- Imbens GW (2020) Potential outcome and directed acyclic graph approaches to causality: relevance for empirical practice in economics. *Journal of Economic Literature* 58, 1129–79. doi: 10.1257/jel.20191597
- Kaaber N (2020) Women's empowerment and economic development: a feminist critique of

- storytelling practices in 'rando-mista' economics. *Feminist Economics* 26, 1–26. doi: 10.1080/13545701.2020.1743338
- King G (1995) Replication, replication. *PS: Political Science and Politics* 28, 444–452.
- King G, Keohane RO and Verba S (1994) *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- King G, Keohane RO and Verba S (2004) The importance of research design. In Brady HE and Collier D (eds), *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Plymouth, UK: Rowman and Littlefield, pp. 111–123.
- Laver M and Sergenti E (2011) *Party Competition: An Agent-Based Model*. Princeton: Princeton University Press.
- Levy JS (2007) Qualitative methods and cross-method dialogue in political science. *Comparative Political Studies* 40, 196–214. doi:10.1177/0010414006296348
- Lieberman ES (2005) Nested analysis as a mixed-method strategy for comparative research. *American Political Science Review* 99, 435–452. doi:10.1017/S0003055405051762
- Mahoney J and Goertz G (2006) A tale of two cultures: contrasting quantitative and qualitative research. *Political Analysis* 14, 227–249. doi:10.1093/pan/mpj017
- Martelli P (2009) Narrazioni analitiche. *Quaderni di Scienza Politica* 3, 553–566.
- Martini S and Olmastroni F (2021) From the lab to the poll: The use of survey experiments in political research. *Italian Political Science Review/Rivista Italiana di Scienza Politica*, 1–19. <https://doi.org/10.1017/ipo.2021.20>
- McCarty N and Meirowitz A (2007) *Political Game Theory: An Introduction*. Cambridge, UK: Cambridge University Press.
- Onwuegbuzie AJ, Johnson RB and Collins KMT (2009) Call for mixed analysis: a philosophical framework for combining qualitative and quantitative approaches. *International Journal of Mixed Research Approaches* 3, 114–139. doi:10.5172/mra.3.2.114
- Pearl J (2009) Causal inference in statistics: an overview. *Statistics Surveys* 3, 96–146. doi:10.1214/09-SS057
- Pemberton J and Cartwright N (2014) *Ceteris paribus* laws need machines to generate them. *Erkenntnis* 79, 1745–1758. doi:10.1007/s10670-014-9639-4
- Ragin CC (2008) *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago, IL: University of Chicago Press.
- Rohlfing I (2008) What you see and what you get: Pitfalls and principles of nested analysis in comparative research. *Comparative Political Studies* 41, 1492–1514.
- Rohlfing I (2012). *Case Studies and Causal Inference: An Integrative Framework*. Basingstoke, UK: Palgrave Macmillan.
- Ruffa C and Evangelista M (2021) Searching for a middle ground? A spectrum of views of causality in qualitative research. *Italian Political Science Review/Rivista Italiana di Scienza Politica*, 1–18. <https://doi.org/10.1017/ipo.2021.10>
- Schmitter P (2008). The design of social and political research. In Della Porta D and Keating M (eds), *Approaches and Methodologies in the Social Sciences A Pluralist Perspective*. Cambridge, UK: Cambridge University Press, pp. 264–295.
- Toshkov D (2016) *Research Designs in Political Science*. London: Palgrave.
- Valentim V, Ruipérez Núñez A and Dinas E (2021) Regression discontinuity designs: a hands-on guide for practice. *Italian Political Science Review/Rivista Italiana di Scienza Politica*, 1–19. <https://doi.org/10.1017/ipo.2021.27>