Chapter 1
Introduction: The Elephant of Causation and the Blind Sages

Alessia Damonte and Fedra Negri

It was six men of Indostan, To learning much inclined, Who went to see the Elephant (Though all of them were blind), That each by observation Might satisfy his mind. John G. Saxe (1816–1887).

Abstract What does a policy outcome hinge on? The response is vital to policy-making and calls for the best of our knowledge from a variety of disciplines—from economics to sociology and from political science to public administration and management. The response entails a stance about causation, however, and almost every discipline has its own. Researchers are like the blind sages who had never come across the elephant of causation before and who develop their idea of the elephant by "touching" a different part of it. Which part of the elephant will you happen to touch? Will you be able to listen to and understand what the other sages will tell you?

1.1 Policy Decisions and Causal Theories

The common wisdom about public policy understands them as governments' decisions to tackle a collective problem. These decisions deploy rules, information, taxes, and expenditures to get "people to do things that they might not otherwise do" or "do things that they might not have done otherwise" (Schneider & Ingram, 1990: 513). By inducing a change in people's willingness and capacity to "do things," policy-makers expect the problem to disappear or, at least, take a more bearable shape.

A. Damonte (✉)
University of Milan, Milan, Italy
e-mail: alessia.damonte@unimi.it

F. Negri
University of Milan - Bicocca, Milan, Italy
e-mail: fedra.negri@unimib.it

A. Damonte, F. Negri (eds.), *Causality in Policy Studies*, Texts in Quantitative
Political Analysis, https://doi.org/10.1007/978-3-031-12982-7_1

Thus, the kernel of policy decisions is the causal theory that they encapsulate: first, of the behavior at the root of the collective problem; second and relatedly, of the capacity that certain tools have to make such behavior change for the better. The theory connects outcomes to behavior and then identifies the "carrots, sticks, and sermons" (Vedung, 2010) best suited to put or keep such behavior on a desirable track. For example, in their fight against cancer, governments can address smoking as a proven causal factor and assume people smoke if they have the wrong information or are shortsighted about the consequences of their behavior—else, they would reasonably quit. Governments can fund education campaigns to convey the right information, require tobacco products to carry warning labels, or disallow tobacco advertising and sponsorship. Moreover, to compensate for people's shortsightedness, they can levy "sin taxes" upon tobacco products to make prices a better signal of the hidden costs of smoking or enforce smoke bans that protect non-smokers. Whether a government applies none, one, or a mix of these tools, in turn, depends on policy-makers; whether their decisions reach the addressees properly, instead, is an administrative and a governance matter (e.g., McConnell, 2010). Regardless of the point of attack, the issue of policy success and failure inevitably appeals to causal theories on endowments, concerns, constraints, and incentives accounting for behavior (e.g., Ostrom, 2005).

Policy studies offer exemplary illustrations of the twofold stake of causal theories. First, these theories allow us to make sense of the world. Our bewilderment at some diversity in performance dissolves when we are offered satisfying accounts of relevant behaviors. Second, these theories have straightforward practical implications for individual and collective strategies. If we know which factors compel an event and suppress it, we can change the event's odds by controlling these factors. Then, the driving question remains: how can we get to know these factors well enough to build decisions on them?

1.2 The Elephant of Causation

Across the philosophy of science and social sciences, the responses to this question invite analogies with the blind sages in Saxe's poem (1872), who "prate about an Elephant that / Not one of them has seen."[1] Indeed, actual causation is the complex local production of an outcome and it is hard to identify before it unfolds. The usable knowledge of a causal process pinpoints the key factors of its unfolding that allow us to see it coming in the next instance and, eventually, change its odds (e.g., Craver and Kaplan, 2020). Such knowledge requires criteria to identify the key

[1] The poem tells the story of a group of blind sages who have never come across an elephant before and who learn what the elephant is like by touching it. Each blind sage feels a different part of the elephant's body, but only one part. They then describe the elephant based on their limited experience and "Though each was partly in the right, /And all were in the wrong!" (Saxe, 1872).

causal factors beyond the single case and credibly so. Historically, guidelines for identifying the key causal factors developed along two lines.

1.2.1 Elephants by the Principle

The most enduring guideline for determining the key causal factors before a process unfolds has come from the Aristotelian philosophy of science. There, causation was tracked back to four kinds of principles, known as “material,” “formal,” “efficient,” and “final.” The first two principles capture the structural features of a causal process, namely, its constituent elements and the shape of their arrangement. The latter two refer to agency and locate the key factors in outer stimuli or the drive from inner purposes (e.g., Moravcsik, 1974). The original “doctrine” maintained that adequate responses to any why-question appealed to all the four principles together.

Indeed, convincing accounts still locate actual causation in the interplay of structure and agency, as influential mechanistic perspectives make clear (e.g., Little, 2011; Craver, 2006). More often, current research streams specialize in single principles. For example, the causal role of “material” ascriptive features is a driving concern of gender and minority studies. The generative power of formal arrangements is the core tenet of, for instance, game theories. Studies on expected utility, values, habits, and emotions take heed of the final goals and motivations, providing fundamental assumptions for neo-institutionalist and behavioral approaches of various stripes. Efficient factors are any stimulus, intervention, or treatment that can elicit a response; thus, they are central to theories of policy instruments, regimes, or political communication, among many others.

With some exceptions (e.g., Bache et al., 2012; Kurki, 2006), current theories seldom claim an explicit legacy with the original canon. The doctrine has fallen into disrepute as improperly scientific, because it invoked a metaphysical reason to justify the causal standing of its four principles. The tenet that individuals with similar features, in a similar situation, with similar motivations, under equivalent stimuli did and will behave in similar ways was justified by the belief that all embodied the same metaphysical essence. As Aristotle argued in a seminal fragment, planets do not twinkle because planets are near things, and not twinkling was intrinsic to near things. Thus, the next planet will not twinkle, too, in force of its “near-thingness.”

This line of reasoning easily lends itself to circular arguments that restate general assumptions instead of probing them. As late as 1673, Molière still had reasons to satirize it. In his comedy *The Hypochondriac*, a “docto doctore” explains in dog Latin that opium makes people sleepy because it embodies a “dormitive virtue.” However, the ultimate criticism came from the British Empiricists, who saw in the appeal to essences a mode for preserving beliefs against evidence and a fundamental obstacle to progress and learning.

1.2.2 Elephants by the Rules

The rejection of metaphysical warrants has called for a different ground for causal inference. Whether a reliable connection exists between being a near thing and not twinkling across cases, so the argument goes, it can only be decided empirically.

Yet, causal evidence does not come to us with labels and numbers attached. Assumptions are still needed about the empirical traces that distinguish between relevant and irrelevant causal factors. In Hume's much-quoted words, causally relevant is:

> an object followed by another and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where, if the first object had not been, the second never had existed. (Hume, 1748, Section VII, Part II, §60).

In short, a factor is relevant to an outcome in the single case under two warrants: the association of the two conforms to a *regular* pattern, and it supports *counterfactual* reasoning.

1.2.2.1 Regularity

The regularity warrant—"where all the objects, similar to the first, are followed by objects similar to the second"—renders the empirical footprint of Aristotelian essences without assuming them and builds on the repeated observation of similar occurrences.

All objects sharing the same feature are similar and constitute a distinct class. Regularity, then, is established between objects in different classes—for instance, in the class of "swan" and in the class of "white." It requires that any observation of the first class entails one in the second. When the regularity holds, causal knowledge can be circulated through handy formulae such as "if a swan, then white."

To apply to the next instance, these formulae have to prove faultless, which is hardly the case: classes and gauges are human constructs and can prove too strict or liberal to capture actual causation in the next instance. Hence, regularity holds provisionally only until we meet the black swan that forces a revision of the scope of our regularity tenets.

Regularity may also seem perfect just because we measured two consequences of the same process. These relationships are useful for prediction; however, they do not qualify as causal as they do not grant control over the events' odds as desired in public policy. Indeed, a barometric reading can be relied upon to prepare for extreme weather conditions but does not license the belief that the coming storm can be tamed by forcing the barometer's pointer. Thus, regularity can be a necessary trait of usable knowledge but insufficient to declare the causal standing of a relationship.

1.2.2.2 Counterfactual

The counterfactual—"where, if the first object had not been, the second never had existed"—enters the picture as the additional warrant to establish causal relevance and ideally applies to the factor in the single case independent of regularity. The warrant borrows from the classical rules of argumentation and the indirect proofs in geometric demonstrations; however, it displays an empirical edge. Counterfactuals link causal relevance to evidence that we could compel a change in the second object by manipulating the first.

From the Humean definition, manipulation is usually understood as suppression; more generally, it means switching the observed state of a feature into its opposite. Thus, counterfactual reasoning requires, first, that we imagine the first object with the switched feature and, then, that we can only draw impossible or contradictory conclusions from it (e.g., Levi, 2007). An exemplary illustration comes directly from Hume. Despite his deep skepticism toward the human mind's ability to fully understand causation, he conceded that our intuitions must be somehow right. To justify his claim, he reasoned that had our mind always got causation wrong (switching the feature), then humankind would have long gone extinct (drawing a conclusion), which contrasts with us thriving as a species (showing the conclusion absurd). Such counterfactual criterion improves on the regularity test, as regular non-causal features fail it: as a broken barometer cannot stop a storm, it cannot be recognized as having any causal standing.

However, counterfactuals have their limits, too. First, they cannot be established unless all the plausible alternative causes of the same outcome are ruled out. Hume's argument does not exclude that humankind's evolutionary success instead depends on, for instance, sheer luck—and the unaccounted alternative undermines the cogency of its conclusion. The second and related issue is serious to the point of earning the title of "fundamental problem of causal inference" in some quarters (e.g., Holland, 1988). Unless we cast the same causal process in the same unit with and without the feature of interest, we cannot establish whether switching the feature can change the outcome.

1.3 The Blind Sages' Portrayals as the Book's Blueprint

The criteria to establish causation by regularity and counterfactual evidence seem as straightforward as impossible to meet. Nevertheless, techniques have been developed as strategies to circumvent the Humean paradoxes and provide empirical warrants to the claim of causal relevance. As Little shows in Chap. 2, technical specialization has undermined the dialogue among techniques and their findings. The appeal to regularity, counterfactual, or mechanistic principles has turned into as many ultimate understandings of causation: "laws" and counterfactuals offered a rival ground for experimental practices; mechanisms took distances from both and licensed causal analysis in actual cases only, under consideration that any conclusion about aggregates necessarily entails an unfaithful reduction—in the end, all models are wrong.

However, the possibility of integration remains when techniques commit to three considerations and are consistent with a reasonable scientific realism. First, causation is real, but our best knowledge of it remains a useful approximation. Second, regularity and counterfactuals are epistemic criteria to establish whether portrayals qualify as valid causal accounts; mechanisms are ontological assumptions about single actual elephants instead. Third, the difference between mechanistic description, models, and laws is not of kind but degree: when they address a common slice of the world, they provide a map of it with different details, abstraction, and scope. Under these commitments, techniques can be understood as devices to respond to special questions about the elephant.

1.3.1 Can this Single Factor Make Any Difference?

The family of experimental and quasi-experimental techniques offers the most renowned, successful, and contentious example at once due to the diffusion of randomized controlled trials as the "gold standard" of scientific knowledge production (e.g., Kabeer, 2020; Deaton & Cartwright, 2018; Dawid, 2000). This family shares the consideration that although we cannot observe a counterfactual directly, we can construe credible "twin worlds" and "treat" one so that the feature of interest provides the only difference to which the difference in responses can be ascribed.

As Battistin and Bertoni show in Chap. 3, this strategy keeps the role of causal assumptions to the minimum required by a stimulus-response model: the treatment is a supposedly efficient cause and connected to performance by a function of a specific shape—often, linear—without further details. Unsurprisingly, these techniques are a cornerstone of usable public policy knowledge: they can establish the capacity of a change in taxation, expenditure, information, and regulation to elicit some effect of interest, apparently without the need for further knowledge.

The credibility of this strategy's conclusions, however, rests heavily on the research design: findings are sound if the twin worlds are construed as statistically identical and independent aggregates, the treatment is forced evenly onto all the units of one world only, and the difference in responses is not affected by the treating procedure or unrelated endogenous dynamics. The threats arise as the statistical aggregates with identical parameters can hide a remarkable inner heterogeneity that may bias both groups' responses in unknown directions. As elaborated by Negri in Chap. 4 and Ornstein in Chap. 5, within the family, this heterogeneity is addressed as the result of selection biases that can be reduced by accounting for observed imbalances and crafting "populations of twins." The solution, however, leaves the issue open of the bending effects from unobservable factors.

The (quasi-)experimental family, in short, can provide reliable measures of the net effect of a treatment, but necessarily at the cost of disregarding the reasons for the diversity in the responses of the treated.

1.3.2 Through Which Structures?

The diversity in responses is instead the driving concern of the second group of techniques. They address it by flipping the experimental balance of model and design and committing themselves to additional assumptions. They conceive of the generative process as patterns of dependence and assign causal relevance to the bundle of factors that fit them.

The reliance on models sidelines the issue of unit selection as, ideally, any unit carries usable information about the tenability of the causal structure of interest. The structure, moreover, provides the fixed points that still make counterfactuals observable. However, models require criteria to select meaningful variables, and structural assumptions provide partial guidance to it. The main decisions can only be made in light of substantive theories about the generation of the outcome—hence, of some previous local knowledge. Within this framework, each technique relies on different languages and pursues different goals.

Path analysis develops within a Bayesian mindset and understands causation as ordered dependencies fitting a few known shapes: chains, colliders, and forks. As Röth clarifies in Chap. 6, these shapes explain because they elaborate on the connection between an alleged causal condition and the dependent by displaying the intermediate causal link, the common factor, or the equivalent alternative factors that support the hypothesis about the unfolding of the causal process before the outcome. The technique supports a neater identification of the mechanism linking a factor of interest and its outcome, affords counterfactual analysis, and provides specific suggestions about the "scope conditions" ensuring the mechanisms. Röth contends that these features qualify path analysis as the natural companion of experimental studies for its capacity to establish the contextual requirements that enhance and refine the validity of their findings.

Qualitative comparative analysis (QCA) instead builds on sets and Boolean algebra and understands causal structures as teams of individually necessary and jointly sufficient factors to an outcome. In Chap. 7, Damonte makes three points about the explanatory import of the technique. First, its assumptions about the shape of causation support complex causal theories about the interactions of triggering, enabling, or shielding conditions of some underlying causal process. Second, its parameters of fit allow diagnosing the underspecification of the theory to the cases at hand, while the algorithm provides a pruning counterfactual device that takes care of its overspecification. Last, sets remap qualities onto quantities, which warrant meaningful and sound solutions. Thus, QCA can formalize and test theories about the teams of conditions beneath policy success and failure across given cases beyond special processes. As such, the technique especially suits the purpose of systematic *ex-post* evaluation of policy designs.

1.3.3 Through Which Process?

The knowledge of the dynamics of a causal situation is the missing piece of knowledge and the core concern of two further strategies, aiming to open up the black box of causation. Both share the direct interest in the actors and their interplay as the ultimate ground of causation, although their point of attack within the causal stream of actions is different.

Bayesian process tracing addresses causation within its local context. In Chap. 8, Bennett shows how analysts can rely on this technique to make causal sense of the chain of events to policy success or failure retrospectively. The strategy understands hypotheses as plausible Bayesian beliefs that we can entertain about the causal process and that evidence can confirm or disconfirm. The weight of evidence rests on the assumption that each hypothesis corresponds to a specific sequence of actions and events that leave empirical traces. When the connection between a piece of evidence and a hypothesis is unique, certain, or both, the actual retrieval of certain traces in a case contributes to ranking hypotheses by their relative likelihood and eventually licenses the ascription of the case to the hypothesis with the best standing.

Last but not the least, agent-based models make it possible to test hypotheses about causal processes as emergent phenomena in silico. As Squazzoni and Bianchi illustrate in Chap. 9, the technique relies on simulation to verify whether a certain alignment of assumptions about actors and their constraints, when translated into conditional rules of individual behavior and recursively played, returns performance values close to the empirical responses of actual systems. The strategy requires regularity and counterfactual assumptions about the options available to each agent, rendered as alternative states, and about the consequence of choosing a state conditional on the states of the relevant neighbors. These models shed light on the tenability of different understandings of the mechanism that alternative policy constraints or endowments activate in the field.

1.3.4 Considerations and Extensions

The order of the chapters, as Beach and Siewert reason in their Chap. 10, chimes with the common prescription in mixed method research that a better causal knowledge follows from a succession of techniques zooming into individual cases, where causation unfolds as actual processes and explanations can find their ultimate validation. However, they consider the downward path of mixed methods lays knowledge open to heterogeneity threats. The actual heterogeneity is always equal to the number of instances under analysis; cross-case knowledge, however, requires that we dismiss some heterogeneity as irrelevant to afford comparisons and causal inferences. The move to local contexts implies a twofold shift—from a low to a high number of factors in the analysis and from coarse types to fine-grained tokens of evidence—that seldom support cross-case findings. Hence, they contend that a more fruitful and conventional strategy follows the upward path from local processes over structures to the causal capacity of single triggers. This path allows more conscious decisions about heterogeneity that can improve models and gauges.

In Chap. 11, Damonte and Negri conclude the journey. The chapter recognizes the fragmented image of causation that the previous contributions convey and asks whether such fragmentation is an undesirable state of affairs, as claimed by a long-honored narrative from the history of science, or an eventually valuable situation, as argued in the pluralist quarters of the philosophy of science. The point of contention concerns the inability to yield dovetailing knowledge that would affect strategies built on alternative tenets. The chapter revises these tenets and contends that, whereas ontology offers complementary angles of attack to the causal elephant and epistemology licenses interpretations that can estrange research communities from one another, methodological reasoning about models and designs reconciles the analyses when it emphasizes that causation corresponds to a few recognized shapes. These shapes, the chapter concludes, offer a rough yet common map of the elephant that strategies of any stripe can detail and enrich while pursuing their special research interests—thus contributing to better policy knowledge.

References

Bache, I., Bulmer, S., & Gunay, D. (2012). Europeanization: A critical realist perspective. In T. Exadaktylos & C. M. Radaell (Eds.), *Research design in European studies* (pp. 64–84). Springer.

Craver, C. F. (2006). When mechanistic models explain. *Synthese, 153*(3), 355–376. https://doi.org/10.1007/s11229-006-9097-x

Craver, C. F., & Kaplan, D. M. (2020). Are more details better? On the norms of completeness for mechanistic explanations. *The British Journal for the Philosophy of Science, 71*(1), 287–319. https://doi.org/10.1093/bjps/axy015

Dawid, A. P. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Association, 95*(450), 407–424. https://doi.org/10.1080/01621459.2000.10474210

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine, 210*, 2–21.

Holland, P. W. (1988). Causal inference path analysis and recursive structural equations models. *ETS Research Report Series, 1988*(1), i–50. https://doi.org/10.1002/j.2330-8516.1988.tb00270.x

Hume, D. (1748). *An enquiry concerning human understanding*. Section VII.

Kabeer, N. (2020). Women's empowerment and economic development: A feminist critique of storytelling practices in 'Randomista' economics. *Feminist Economics, 26*(2), 1–26. https://doi.org/10.1080/13545701.2020.1743338

Kurki, M. (2006). Causes of a divided discipline: Rethinking the concept of cause in international relations theory. *Review of International Studies, 32*(2), 189–216. https://doi.org/10.1017/s026021050600698x

Levi, I. (2007). *For the sake of the argument: Ramsey test conditionals, inductive inference and nonmonotonic reasoning*. Cambridge University Press.

Little, D. (2011). Causal mechanisms in the social realm. In P. M. K. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 273–295). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199574131.003.0013

McConnell, A. (2010). Policy success, policy failure and gray areas in-between. *Journal of Public Policy, 30*(3), 345–362. https://doi.org/10.1017/S0143814X10000152

Moravcsik, J. M. E. (1974). Aristotle on adequate explanations. *Synthese, 28*, 3–17. https://doi.org/10.1007/BF00869493

Ostrom, E. (2005). *Understanding institutional diversity*. Princeton University Press.

Schneider, A., & Ingram, H. (1990). Behavioral assumptions of policy tools. *The Journal of Politics, 52*(2), 510–529. https://doi.org/10.2307/2131904

Vedung, E. (2010). Policy instruments: Typologies and theories. In M.-L. Bemelmans-Videc, R. C. Rist, & E. Vedung (Eds.), *Carrots, sticks and sermons: Policy instruments and their evaluation* (pp. 21–58). Transaction. https://doi.org/10.4324/9781315081748