Causation

Michael Baumgartner

Abstract

Although the concept of causation is ubiquitous both in science and in every-day life, it is far from pre-theoretically clear. Theories of causation are designed to render precise what causation amounts to and, thereby, to provide conceptual frameworks against the background of which empirical research into causation becomes possible. This article, first, discusses the dimensions of variance between different theories and, second, presents the core ideas behind those theories with highest relevance for political science: the regularity theory, the probabilistic theory, the counterfactual theory, the interventionist theory, and the mechanistic theory.

1 A history of disagreement

Causation is one of the most basic concepts regulating our interaction with the world. It is omnipresent both in science and in every-day life. Correspondingly, it is among the oldest topics in western philosophical and scientific theorizing. Aristoteles, Aquinas, Descartes, Hobbes, Galileo, Spinoza, Leibniz, Locke, Newton, Hume, Kant, Mill, Reichenbach—just to name some of the most prominent figures—have devoted important parts of their work to this topic. A multitude of different theories of causation, many of which incompatible, have been proposed over the centuries. Even to this day, conflicting theories continue to co-exist—ultimately because they are embedded in, and draw their justification from, incompatible background metaphysics and ontologies, which are notoriously difficult to reconcile (see also the chapter by Moses in this Handbook). In consequence, despite in-depth and centuries-old investigations into the topic no consensus has been reached on the most fundamental question: what is causation? Is it an objective feature of our world or is it something we, as observers, project onto the world? Is it something that actually governs what occurs around us or is it a concept that merely facilitates theorizing about those occurrences? Is it a matter of the instantiation of regularities or laws, or of counterfactual dependence, or of probability raising, or of manipulability, or of mechanisms? Does it only obtain between occurrences in space and time or does it also obtain between absences of such occurrences? And more specifically, is it a transitive relation or not; is it a deterministic relation or not?

It is beyond the scope of this article to give an overview over the multitude of arguments for and against the different answers that have been given to these questions over time. Instead, I focus on establishing the importance of developing explicit theories of causation rendering transparent the understanding of causation presupposed in a given research context. Moreover, the article provides

systematic introductions to those theories with most conceptual and methodological impact for political science: the regularity theory, the probabilistic theory, the counterfactual theory, the interventionist theory, and the mechanistic theory. Where appropriate, the historical roots of these theories will be discussed, but my focus will be on their modern versions. Readers interested in an introduction to the topic with a historical focus are referred to Bunge (1979). An overview of the history of causation is provided in Beebee et al. (2009, part I). The article ends with an outlook identifying methodological frameworks suitable for uncovering causation as defined by the discussed theories.

2 Purpose of a theory of causation

Despite the omnipresence of causation in every-day life and science, it is far from pre-theoretically (i.e. intuitively) clear under what conditions a causal relation obtains. Even in the most commonplace of scenarios, it easily happens that our causal judgements are unclear, unstable, and inconsistent. To illustrate, consider Case 1.

Case 1. *Angelo lives in Rome and goes on vacation. His neighbor agrees to water Angelo's plants, but repeatedly forgets to do so. When Angelo returns two weeks later, his plants are dead.*

When asked to identify the causes of the plants' death, most people will, presumably, point to the neighbor's negligence and contend that the ultimate cause of death was insufficient water supply. Of course, insufficient water supply did not only result from the neighbor's negligence but also from the fact that everybody else, including, say, the Pope did not water regularly. Yet, even though it is uncontroversial that, had the Pope watered the plants, they would have survived, many people will deny that the Pope's inaction is another cause of death. How are these discrepancies in causal relevance ascription justifiable? One important difference between the neighbor and the Pope is that the former has explicitly agreed to water the plants, whereas the latter has not. Yet, while the violation of the agreement is a reason to hold the neighbor and not the Pope morally accountable, it does not seem to be a reason to ascribe causal relevance to the neighbor's but not the Pope's inaction. After all, either one could have secured sufficient water supply; and causation is a relation that obtains in the world independently of judgements of moral accountability—or so it seems. But then, either the neighbor's and the Pope's (and everybody else's) inactivity caused the plants to die or neither of them did.

The list of unclear candidate causes connected to Case 1 can easily be extended. Is Angelo's failure to give the plants to his (reliable) mother prior to departure also a cause of their death? Or what about Angelo's purchase of the plants; or Angelo's birth? There is a case to be made that, had he not been born, he would not have bought the plants, which, subsequently, would have been bought by some

plant enthusiast. I leave it to the reader to contemplate possible answers to these questions. What matters for our purposes is that, even though Case 1 describes a most commonplace scenario, causally analyzing it in a clear and consistent manner is far from trivial. And the difficulties are not due to lacking empirical evidence. We can safely assume that we have enough evidence on the neighbor's, the Pope's, and the mother's watering behaviors, on how water influences plant growth, and on possible alternative plant buyers etc. to determine exactly what would have happened in relevant contrast scenarios. That does not help us to determine, for example, whether we should ascribe causal relevance to inaction, or whether we should take causation to be a transitive relation and count Angelo's birth among the causes of the plants' death. These, and many others, are *conceptual decisions* that need to be taken before the question what the available evidence entails for causal relations even arises. It is the purpose of a theory of causation to take those conceptual decisions, render them transparent in an explicit definition of causation and, thereby, provide a framework against the background of which empirical research into thus defined causation becomes possible.

A theory of causation provides necessary and sufficient conditions for a dependency to be of causal nature, or differently, it specifies the truth conditions of claims of the form 'x causes y'. This is accomplished by replacing P with suitable analysis conditions in schema (Φ), where "iff" is short for "if, and only if":

$$x \text{ causes } y \text{ iff } P. \qquad (\Phi)$$

Causation is not a technical term of art that can be stipulatively defined in any way we please, rather, it is widely used in every-day and scientific language. A theory that wants to be taken seriously must take that existing usage into account and aim for a P that reproduces standardly accepted pre-theoretic causal judgements. However, as these judgements tend to be inconsistent, no (consistent) theory can possibly reproduce all of them. At best, a theory can do justice to a maximally large consistent proper subset of pre-theoretic causal judgements. The selection of these subsets can vary from theory to theory and there is no fact of the matter as to what is the true or even best selection. Rather, the selection must be justified based on pragmatic considerations concerning, for instance, the theoretical or methodological purposes a particular theory is intended to serve. In that light, the position of *causal pluralism*, that is, the view that different theories of causation do not contradict one another but capture different concepts or variants of causation, has seen a rise in popularity in recent years (e.g. Psillos 2010). Correspondingly, which theory to choose in a given research context depends on the exigencies of that context, which are determined, for instance, by the investigated research questions or by the nature of the available data.

3 Dimensions of variance between different theories

Theories of causation differ along various dimensions. This section discusses the most relevant ones.

Reductionism vs. non-reductionism

All theorizing about the world—in any domain—needs to start with some conceptual inventory that is presupposed as being clear. The concepts in that inventory are called *fundamental*. The most general divide between different theories of causation concerns the question whether the concept of causation should be considered one of those fundamental ones or not. A theory, subject to which causation is fundamental, holds that it is impossible to define causation without recourse to some 'causally loaded' concept or other, meaning that it is impossible to substitute P in schema (Φ) by conditions that are entirely free of causal connotations. Such a theory is called *non-reductionist*, for it contends that it is impossible to reduce causation to non-causation. By contrast, a *reductionist* theory maintains that causation is not fundamental, that is, causation can be defined in terms of entirely non-causal concepts—it can be reduced to non-causation. Hence, a reductionist theory substitutes P by conditions without any causal connotations whatsoever.

To illustrate, an interventionist theory is non-reductionist, for it defines causation in terms of interventions, which are specific types of causes. The idea is that certain causes, *viz.* interventions, are fundamental and pre-theoretically identifiable and can be used to identify all non-fundamental causes. By contrast, a regularity theory is reductionist, as it defines causation in terms of Boolean dependencies of sufficiency and necessity, which are entirely non-causal.

Overall, the divide between reductionism and non-reductionism partitions the currently available theories of causation in roughly equal halves.

The ontology of causation

The second dimension of variance concerns the ontology of causation, that is, the question what types of entities can stand in causal relations. The answers to that question are influenced by two conflicting intuitions. On the one hand, it seems that causes and effects are entities that *occur* in time and space. For example, elections are causes of policy changes or peasant revolts are causes of social revolutions. Causes and their effects are items we can point to. On the other hand, absences and omissions are often causally interpreted as well, even though their characteristic feature is that they *do not occur* in time and space. For instance, the absence of water causes plants to die or the absence of political participation causes popular anger.

Corresponding to these two intuitions, there are two main candidate ontologies of causation: according to the first, causes and effects are *events* (e.g. Davidson 1967), according to the second they are *facts* (e.g. Mellor 1995). There exist many elaborate theories of both events and facts, which cannot be reviewed here (cf. Casati and Varzi 2015). What matters for our purposes is that events and facts constitute categorically different entities: events are spatiotemporally located concrete entities, facts are non-spatiotemporally located abstract entities. Obama's election as first African-American president is an event that occurred on Novem-

ber 4, 2008, in the USA, whereas the fact that Obama was elected as the first African-American president is not located in the year 2008 in the USA, rather it is a fact in contemporary Europe just as it will be a fact in Australia 100 years from now.

The categorical difference between events and facts yields that event and fact ontologies cannot be combined in one theory of causation. Either a theory takes causes and effects to be events or facts (not both). Yet, neither ontology is clearly preferable over the other; there are persuasive arguments for and against both ontologies (for an overview cf. Ehring 2010). For that reason it has become customary to bracket the question as to the ontology of causation by remaining as non-committal as possible with regard to the nature of causes and effects. This is accomplished by referring to causes and effects simply as *variables*, *factors*, or *values* of variables/factors. Random variables and factors are flexible modeling devices that can be used to represent any types of entities. On the upside, theories of causation that are formulated in terms of variables or factors can, hence, easily be adapted for different ontologies; on the downside, what such theories have to say about causation becomes relative to the choice of analyzed variables/factors, which introduces a subjective element into causal analyses (cf. Halpern and Hitchcock 2010).

General vs. singular causation

Two kinds of causation must be distinguished: one on the type level, *general* or *type* causation (often also called *causal relevance*), and one on the token level, *singular* or *token* causation (often also called *actual* causation). Examples of the first kind are 'regular watering causes plant growth' or 'printing money causes inflation'; examples of the second are 'Turkey's money printing in 2018 causes Turkey's inflation in 2018' or 'the Titanic's collision with an iceberg on 15 April, 1912, caused the Titanic's sinking'. General causation relates types (of events or facts) that can be instantiated repeatedly and on various occasions by corresponding tokens, which themselves are related in terms of singular causation. General causation is the kind of causation that is commonly traced in scientific theory-building and causal modeling and that figures in scientific laws. Knowledge of general causation is needed for prediction. Singular causation, by contrast, is commonly traced in the course of explaining some event or fact of interest. Knowledge of singular causation is needed for retrodiction.

Plainly, general and singular causation are not independent. If some type A is causally relevant to another type B, there exists a token α of type A (in the right circumstances) that is a singular cause of a token β of type B; and if a token α is a singular cause of another token β, there exists a type-level structure featuring the corresponding type A as general cause of B. Accordingly, theories of general and singular causation are not independent either. The standard approach in theory development, therefore, is to first define either general or singular causation—as so-called *primary analysandum*—and to then spell out the other kind of causation

in terms of the primary analysandum. As we shall see below, some theories take general causation to be primary, others opt for singular causation.

Relational properties of causation

Causation, both on type and on token level, is usually analyzed as a relation '... is causally relevant to ...' or '... is a singular cause of ...', where the dots are filled by the corresponding causes and effects. Every relation can be characterized by the following relational properties: symmetry, reflexivity, and transitivity. To properly understand a relation clarifying its relational properties is crucial.

Some relational properties of causation are uncontroversial, others not. For instance, general and singular causation are clearly *neither reflexive nor symmetric*. That is, a cause is normally neither caused by its own effects nor by itself. Standardly, it is even assumed that, on the token level, there are no causal feedbacks at all. Token causes and effects are distinct entities such that the former temporally precede the latter. The tax raise at time t_1 causes firms to lay-off people at a later time t_2, but at t_2 the tax raise has already happened, such that the lay-offs cannot cause the initial tax raise. However, feedbacks may occur on the type-level. Tax raises may be causally relevant to lay-offs, which cause the need to generate more revenue for social security, which, in turn, is causally relevant to further tax raises. That is, on the type level it is possible that a cause, via a sequence of intermediate links, is causally relevant to itself, thereby closing a causal cycle.

Concerning the third relational property, transitivity, matters are much less clear. While it certainly often holds that a cause A not only brings about its direct effect B but also (indirectly) causes B's effects, the question as to the transitivity of causation is whether causal influence is *always* propagated along causal chains. Is it generally the case that, if A is cause of B and B is a cause of C, then A is also a cause of C? Here again, there are conflicting intuitions. On the one hand, more often than not, our objectives when interacting with the world cannot be caused directly by one suitable intervention; rather, our ultimate objectives are typically far remote from what we can cause directly. This holds in particular in politics. If our objective is, say, to reduce the unemployment rate, all we can do is to bring certain monetary, fiscal or educational policy changes under way, which, via many intermediate links, may ultimately reduce unemployment. We would not induce these policy changes (with often unwanted side-effects) if we were not confident that they will ultimately cause the desired objective. Hence, much of our interaction with the world is driven by the assumption that we can rely on our actions not only having direct effects but also far remote indirect ones, which amounts to the assumption that causation is transitive.

Yet, in the face of certain concrete examples, the intuition that causation is transitive becomes shaky. To illustrate, consider Case 2.

Case 2. *Walter is a candidate in a presidential race. But his poll numbers are going down. In a nationally televised debate he intensifies his populist campaign pledges. In the next poll, his numbers are going up again.*

Walter fuels his populism as a reaction to his sinking polls, and populism tends to be conducive to rising polls. That is, there is a causal chain from sinking polls to intensified populism and on to rising polls. If causation is transitive, it follows that sinking polls cause rising polls, which seems highly counterintuitive. There are some authors arguing that transitivity can be maintained even in light of examples as Case 2 (e.g. Lewis 2000), others contend that such examples prove that intuitions as to the transitivity of causation are misguided (e.g. Hitchcock 2001). Correspondingly, there are theories of causation providing transitive notions of causation, while others define causation in such a way that transitivity does not hold.

Realism vs. anti-realism

Another dimension of variance between theories of causation is their stance on the question whether the causal relation is a real constituent of the world. While it is beyond doubt that causes, effects, and their behavior patterns exist, it is a matter of controversy whether there additionally exists a causal bond connecting them and governing their behavior. The position of *causal realism* maintains that not only the entities related by the causal relation exist, but also the relation itself; causes and effects are connected by a real causal bond (e.g. Tooley 1987). By contrast, the position of *causal anti-realism* contends that only the entities related by the causal relation and their behavior patterns exist, not the relation itself; there are no causal bonds (Hume 1748). Certain theories of causation endorse causal realism, others subscribe to causal anti-realism.

According to an anti-realist theory, causation boils down to the regularities, the correlations, or other sorts of dependencies obtaining between the behaviors of causes and effects. To be causally related is nothing over and above behaving in a certain (yet to be specified) manner. Subject to a realist theory, by contrast, causation is more than a certain behavior pattern. To be causally related means to be tied together by a causal bond.

The difference between realism and anti-realism is not just of philosophical interest. The stance taken on this issue has a direct bearing on what can count as empirical evidence for causation. To establish causation, the anti-realist only has to demonstrate the existence of the required behavior pattern in the data—and behavior patterns are exactly what is contained in ordinary data. By contrast, the realist has to furnish evidence for the existence of the relevant type of causal bond. But contrary to behavior patterns, such bonds are not directly visible or measurable, meaning that ordinary data will not contain direct evidence on causal bonds. Instead, the existence of bonds must be indirectly inferred from the (behavioral) information contained in data.

In light of the fact that causation as defined by an anti-realist theory is more directly accessible in data, the theories with most relevance in empirical science are of the anti-realist type. The only theory with realist leanings resorted to in political science is the mechanist theory.

Production vs. difference-making

The contrast between realist and anti-realist theories closely aligns with the contrast between so-called *production* and *difference-making* theories. On the one hand, a production theory contends that the characteristic feature of causes is their capacity or disposition to produce the effect, where production is taken to be a fundamental (i.e. non-reducible) causal notion expressing the physical bringing about of an effect, for instance, via the transfer of energy or momentum (e.g. Dowe 2000). On the other hand, a difference-making theory maintains that the characteristic feature of causes is that they are difference-makers of their effects, which is to be understood in terms of the non-causal notion of association: A is a difference-maker of B iff there (possibly) exist homogenous scenarios in which a difference in A is associated with a difference in B (e.g. Woodward 2003).

Production theories tend to subscribe to realism, difference-making theories to anti-realism. Correspondingly, causation as defined in difference-making terms can be more easily tracked in data. While difference-making relations can likewise be investigated on coarse-grained macro and fine-grained micro levels, tracing production relations requires zooming in on the micro-level in order to determine what is happening between causes and effects.[1]

Determinism

Until the development of quantum mechanics in the early the 20^{th} century, causation was generally believed to satisfy the *principle of determinism*: whenever the same types of causes occur, the same types of effects occur as well; in slogan form, 'same causes, same effects'. Of course it had always been recognized that determinism is often not visible in data. For example, although regular watering causes plant growth, some regularly watered plants die nonetheless. However, such indeterminacies were taken to be a result of insufficient control over background influences generating noise, meaning that scenarios with seemingly alike causes but different effects do not actually feature the exact same causes. Our world is of such enormous causal complexity that it is often impossible to ensure that nothing interferes with a cause or that all background conditions necessary for a cause to be efficacious are actually instantiated. But other things being equal, *ceteris paribus*, causation is a deterministic dependence relation.

However, various so-called *no-hidden-variables* theorems in quantum mechanics suggest that indeterminacies in data are not always due to noise but—at least on the level of fundamental particles—a result of the inherent indeterministic nature of the physical processes themselves (Albert 1992). Hence, according to the standard interpretation of the mathematical machinery of quantum mechanics, the principle of determinism is false. Yet, even though quantum mechanics is

[1] As studies on multiple levels of granularity are not mutually exclusive, Russo and Williamson (2007) have influentially argued that, ultimately, it takes both difference-making and production evidence to establish causation.

one of the most successful scientific theories currently at our disposal, the indeterministic interpretation of its mathematical formalism failed to convince many friends of determinism that the principle of determinism must be abandoned—for two main reasons. On the one hand, there also exist deterministic interpretations of the quantum mechanical machinery, e.g. Bohm's interpretation (Albert 1992, ch. 7). On the other hand, even if it should turn out that there exist inherently indeterministic processes on the fundamental level, there are many open questions with respect to the *causal* interpretability of these processes (cf. e.g. Healey 2010).

Therefore, quantum mechanics notwithstanding, many currently available theories of causation either explicitly endorse the principle of determinism or remain non-committal as to its validity. The possibility of fundamental indeterminism induced by quantum mechanics has been the main motivation behind the development of probabilistic theories of causation.

4 Main theories

This section reviews the theories of causation with highest relevance for social and political science.

Regularity theory

So-called *regularity theories* of causation have the longest tradition among theoretical frameworks that continue to be developed today. Their development dates back to Hume (1748). The most influential modern regularity theory is due to Mackie (1974)—with refinements by Graßhoff and May (2001) and Baumgartner (2013). In regard to the theoretical dimensions of variance, regularity theories make the following analytical choices. Their primary analysandum is general causation, which they reductively define in terms of redundancy-free regularities obtaining among factors taking on specific values, such as 'whenever factor A takes value i (A=i), factor B takes value j (B=j)'. Moreover, they subscribe to anti-realism and assume that causation is a deterministic form of dependence that is not transitive. Finally, they contend that the characteristic feature of causes is that they are difference-makers of their effects.

Factors analyzed by regularity theories can either be *crisp-set*, taking two possible values 0 and 1, *fuzzy-set*, taking continuous values from the unit interval $[0, 1]$, or *multi-value*, taking an open (but finite) number of possible values $\{0, 1, 2, \ldots, n\}$. For simplicity of exposition, I focus on crisp-set factors here, which allows for conveniently abbreviating the explicit 'Factor=value' notation. As is conventional in Boolean algebra, 'A' is short for A=1 and 'a' for A=0. Modern regularity theories moreover borrow much of the formal machinery from Boolean algebra, in particular, the operations of *conjunction*, $A{*}B$ (expressing 'A=1 and B=1'), *disjunction*, $A + B$ ('A=1 or B=1'), *implication*, $A \rightarrow B$ ('if A=1, then B=1'), and *equivalence* $A \leftrightarrow B$ ('A=1 if, and only if, B=1') (see also Wagemann in this Handbook).

The implication operator allows for formally expressing regularities, more specifically, it allows for defining the notions of *sufficiency* and *necessity*, which are the two core Boolean dependencies exploited by regularity theories: A is sufficient for B iff $A \rightarrow B$ (i.e. whenever A is given, B is given); A is necessary for B iff $B \rightarrow A$ (i.e. whenever B is given, A is given). Many of these Boolean dependencies, however, have nothing to do with causation. For example, the sinking of a (properly functioning) barometer is sufficient for weather changes but it does not cause the weather; or whenever there is an election, votes are cast, so casting votes is necessary for the election—but it does not cause the election. Still, some Boolean dependencies are in fact due to underlying causal dependencies: rainfall is sufficient for wet streets and also a cause thereof, or winning an election is necessary for being sworn into office and also a cause thereof.

That means the crucial problem to be solved by a regularity theory is to filter out those Boolean dependencies that are due to underlying causal dependencies and are, hence, amenable to a causal interpretation. The main reason why most structures of Boolean dependencies do not reflect causation is that they tend to contain redundancies, whereas structures of causal dependencies do not feature redundant elements. Every part of a causal structure makes a difference to the behavior of that structure in at least one context. Accordingly, to filter out the causally interpretable Boolean dependencies, they need to be freed of redundancies. Only those elements of sufficient and necessary conditions can be causally relevant which are indispensable to account for a scrutinized outcome in at least one context. Or in Mackie's (1974, 62) words, causes are at least *INUS conditions*, *viz.* insufficient but non-redundant parts of unnecessary but sufficient conditions.

Whatever can be removed from sufficient and necessary conditions without affecting their sufficiency and necessity is not a difference-maker and, hence, not a cause. The causally interesting sufficient and necessary conditions, hence, are *minimal* in the sense that they do not contain sufficient and necessary proper parts, respectively. Minimally sufficient and minimally necessary conditions can be combined in so-called *minimal theories* (Graßhoff and May 2001), which constitute the heart of contemporary regularity theories: a minimal theory of an outcome B is a minimally necessary disjunction of minimally sufficient conditions of B.[2] An example might be:

$$A{*}e \,+\, C{*}d \;\leftrightarrow\; B \tag{1}$$

(1) being a minimal theory of B entails that $A{*}e$ and $C{*}d$, but neither A, e, C, nor d alone, are sufficient for B and that $A{*}e \,+\, C{*}d$, but neither $A{*}e$ nor $C{*}d$ alone, are necessary for B.

Minimal theories directly mirror the complexity of causes (conjunctural causation) and the principle of equifinality: causes do not bring about their effects in isolation but only in conjunction with other causes, and outcomes can be caused

[2]To do justice to the different types of redundancies that Boolean dependency structures may be affected by, the complete definition of the notion of a minimal theory is intricate and beyond the scope of this article (for the latest definition see Baumgartner and Falk 2018).

along various alternative paths. Moreover, although minimal theories as (1) feature a main operator ('$\leftrightarrow$') that is symmetric, to the effect that both sides of (1) are mutually sufficient and necessary for one another, the fact that (1) features *two* minimally sufficient conditions of B yields that it nonetheless identifies a direction of determination: both $A{*}e$ and $C{*}d$ are sufficient for B, but B is neither sufficient for $A{*}e$ nor for $C{*}d$. Hence, $A{*}e$ and $C{*}d$ are the causes of B, and not vice versa.

Still, to define causal relevance by means of minimal theories, an additional constraint is needed, for not all minimal theories faithfully represent causation. In a nutshell, the reason is that complete redundancy elimination is relative to the set of analyzed factors **F**, meaning that factor values contained in minimal theories relative to some **F** may fail to be part of a minimal theory relative to a superset of **F** (Baumgartner 2013). In other words, by expanding sets of analyzed factors, factors values that appear to be non-redundant to account for an outcome can turn out to be redundant after all. Only factor values that are not rendered redundant by expanding factor sets are causally relevant.

These considerations yield the following substitution instance of schema (Φ):

(R) A is a type-level cause of B iff A is part of a minimal theory of B relative to a factor set **F** and remains part of a minimal theory of B across all expansions of **F**.

The main problem of (R) is that, since Boolean dependencies expressed in minimal theories are inherently deterministic, (R) incorporates determinism at its very core. But it may turn out that some of the indeterminism we typically encounter in data is not due to noise but to the fact that some causal relations are inherently indeterministic. Hence, (R) is a viable theory only for contexts where causation can safely be assumed to be deterministic. While that assumption is dubious for certain micro-level areas explored by physics, it seems innocuous for macro-level areas investigated in social and political sciences.

Probabilistic theory

The restriction of regularity theories to deterministic contexts has prompted *probabilistic* theories to abandon the assumption that causation is deterministic (Reichenbach 1956; Suppes 1970). Still, like regularity theories, probabilistic accounts are difference-making theories that commit to causal anti-realism, and they do not entail that causation is transitive. Apart from that, there is much variance within the probabilistic framework. Some probabilistic theories take general causation to be the primary analysandum (Eells 1991), others take singular causation to be primary (Glynn 2011), still others contend that both general and singular causation can be analyzed in one go (Suppes 1970). There are reductionist accounts (Glynn 2011; Suppes 1970), and non-reductionist ones (Cartwright 1979; Eells 1991). In what follows, I focus on theories that primarily analyze general causation and sketch the main ideas behind one reductionist and one non-reductionist account.

Like regularity theories, probabilistic theories remain non-committal with respect to the ontology of causation by referring to causes and effects as variables or factors taking values. For simplicity, I continue to concentrate on the case of binary factors and to employ the Boolean shorthand notation introduced in the previous section. Probabilistic theories reject the regularity theoretic requirement that causes are parts of sufficient conditions of their effects. Instead, A can count as a cause of B if A merely raises the probability of B, where A is said to raise the probability of B iff the probability of B conditional on A is higher than the probability of B conditional on not-A, *viz.* a, or formally:

$$\mathrm{P}(B \mid A) > \mathrm{P}(B \mid a) \tag{2}$$

However, not all cases of probability-raising are also cases of causation, for several reasons. First, probability-raising is symmetric: If A raises the probability of B, then B also raises the probability of A. But causation is not symmetric. There are various ways to address that problem. For instance, in complex networks of probabilistic (in-)dependencies, so-called *Bayesian networks*, it is possible to infer the direction of causation from substructures featuring multiple independent paths to the same effect, so-called *unshielded colliders* (Spirtes et al. 2000, ch. 5). Another way to distinguish between causes and effects is via their temporal order: causes precede their effects. This can be formally captured by time indexing probabilistically related factor values, to the effect that the probability-raiser $A_{t'}$ precedes the raised B_t:

$$\mathrm{P}(B_t \mid A_{t'}) > \mathrm{P}(B_t \mid a_{t'}) \,, \quad \text{where } t' < t \tag{3}$$

Second, many cases of temporally ordered probability-raising in the vein of (3) are not due to a causal dependence between $A_{t'}$ and B_t but to a common cause of $A_{t'}$ and B_t. For example, the sinking of a barometer at t' raises the probability of rain at a later time t without causing it. This probabilistic dependence is the result of an approaching low-pressure system at a time t'' before t' causing the barometer to sink on one path and the rain on another. Accordingly, conditional on the low-pressure system, a sinking barometer no longer raises the probability of rain; in other words, given that a low-pressure system is approaching, additional information about the behavior of a barometer has no bearing on the rain probability. More generally, the set of common causes $\mathbf{C}$ of two parallel effects A and B neutralizes or *screens off* the probabilistic dependence between A and B in the following sense (Reichenbach 1956) (where '$*$' again stands for conjunction):

$$\mathrm{P}(B \mid A * \mathbf{C}) = \mathrm{P}(B \mid a * \mathbf{C}) \tag{4}$$

Relations of temporally ordered probability-raising only track causation if they are not screened off by antecedently instantiated factors. This idea is captured in the following reductionist theory (Suppes 1970):

($\mathbf{P}_1$) $A_{t'}$ is a type-level cause of B_t, where $t' < t$, iff $A_{t'}$ raises the probability of B_t and there does not exist a set $\mathbf{C}_{t''}$, where $t'' < t'$, that screens off the probabilistic dependence between $A_{t'}$ and B_t.

(P_1) has two problems. First, not all causes raise the probability of their effects; some in fact lower it. To cite a classical example, wind gusts lower the probability that golfers make a hole-in-one. Nonetheless, it can happen that wind gusts deflect balls in such a way that they end up in the hole in one shot after all (e.g. by luckily bouncing off trees). Being essential contributors to the trajectories of such balls, the wind gusts are causes of the holes-in-one, even though they lower their probability. So, causes must not be required to be probability-*raisers* of their effects, rather, it suffices that a cause $A_{t'}$ is a probability-*changer* of its effect B_t in the following sense:

$$\mathrm{P}(B_t \,|\, A_{t'}) \neq \mathrm{P}(B_t \,|\, a_{t'}) \;, \quad \text{where } \; t' < t \tag{5}$$

The second problem of (P_1) stems from the fact that probabilities are typically determined via relative frequencies in a studied population. Probabilistic dependencies inferred from relative frequencies in the whole population, however, may be reversed or neutralized in subpopulations, which is a very common phenomenon in statistics known as *Simpson's Paradox*. To cite an example used by Cartwright (1979), it can happen that, in the whole population of applicants to some university X, the frequency of admission among male applicants is significantly higher than among female applicants, while in the subpopulations of X's faculties, the admission rates among men and women are exactly equal (for illustrations see Eells 1991, 62-80). Such frequency distributions could be the result of men more often applying to faculties that are easier to get into. But unequal ratios could reappear in even more fine-grained subpopulations; for instance, it could turn out that men are more frequently admitted than women to each of X's departments. It is thus unclear whether being male should be regarded as a probability-changer of being admitted to university X or not.

Many representatives of probabilistic theories of causation have taken examples of such paradoxical frequency distributions to show that rigorous constraints must be imposed on populations suitable for inferring probability-changing relations, in particular, that such populations must be required to be homogeneous in *causally relevant respects*. Or put differently, a probability-changing relation between $A_{t'}$ and B_t that tracks causation must obtain in a *causal context* $\mathbf{K}$ in which all causes of B_t not on a causal path from $A_{t'}$ to B_t are constant. This leads to the following non-reductionist theory (Eells 1991, 86, 106):

($\mathbf{P}_2$) $A_{t'}$ is a type-level cause of B_t, where $t' < t$, iff there exists a causal context $\mathbf{K}$ such that, in $\mathbf{K}$, $A_{t'}$ changes the probability of B_t and there does not exist a set $\mathbf{C}_{t''}$, where $t'' < t'$, that screens off the probabilistic dependence between $A_{t'}$ and B_t.

(P_2) is non-reductionist because it does not define causal relevance in non-causal terms but in terms of *causal* contexts. Applying (P_2) to concrete cases, say, in the course of identifying the causes of some outcome B, hence, presupposes substantive prior causal knowledge about B's causes. Overall, (P_2) avoids the problems of (P_1), but it does so at the price of abandoning the project of defining causation in non-causal terms.

Counterfactual theory

A theoretical framework not prepared to give up reductionism is the so-called *counterfactual* one. Like regularity theories, counterfactual theories have their roots in suggestions by Hume (1748). They were developed into a full-blown theory by Lewis (1973; 2000). Further refining Lewis' account is a field of ongoing research, incorporating various techniques from structural equation modeling (e.g. Halpern 2016). As the technicality of these latest proposals is beyond the scope of this article, I will subsequently concentrate on Lewis' original theory. It is a reductionist difference-making theory that subscribes to anti-realism and assumes causation to be a deterministic form of dependence. Contrary to the previously discussed theories, it stipulates that causation is transitive, it focuses on singular causation as primary analysandum, and it takes a determinate stance on the ontology of causation by assuming that causes and effects are spatiotemporally located events. In order not to confuse events with factors, I subsequently refer to events using Greek lower case letters α, β, etc.

The main idea behind a counterfactual theory is to define a singular causal relation between two occurring events α and β in terms of the truth of counterfactual conditionals of the form 'had α not occurred, β would not have occurred'. Or more concretely, Turkey's money printing in 2018 is a cause of Turkey's inflation in 2018 if it is true that, had Turkey not printed money in 2018, there would have been no inflation in Turkey in 2018. Even though such counterfactual claims are ubiquitous in every-day language, it is difficult to state precisely under what conditions they are true. The main problem is that they refer to *non-actual* scenarios, meaning their truth cannot be determined by observing, measuring or conducting experiments. Their truth also does not depend on different scenarios of the same type that actually occurred (on different occasions). For example, that Turkey did not print money in 2014 and did not have an inflation in that year does not tell us what would have happened in 2018, had there been no money printing.

To render the truth conditions of counterfactual statements precise, it is standard to draw on so-called *possible world semantics*, which were developed in modal logic to explicate the meanings of claims about possibility and necessity. A possible world **w** can be thought of as a (hypothetical) maximal state of affairs such that every state of affairs is either included in **w** or precluded by **w**. There is a possible world in which McCain (and not Obama) becomes president in 2008, Italy is in Africa (and not in Europe), Caesar (and not Armstrong) is the first man on the moon, Archduke Ferdinand is not assassinated but lives to be 80 years old,

etc. Importantly, all possible worlds can be compared as to how similar or distant they are relative to one another, and there exists one distinguished possible world, the *actual world*, including all the states of affairs that obtain in the world we live in. Moreover, a possible world **w** is said to be an *α-world* if the event α occurs in **w**, and a *non-α-world* otherwise.

Relative to that theoretical background, Lewis (1973) determines that the statement 'had α not occurred, β would not have occurred' (where α and β are events in the actual world) is true iff some non-α-world where β does not occur is more similar to the actual world than any non-α-world where β occurs, or differently, iff it takes less of a departure from actuality to suppress α and β together than to just suppress α. If that counterfactual statement is true, α and β are said to be *counterfactually dependent*, and α is a cause of β.

Before this account can be applied to identifying token-level causes, the conditions under which a non-α-world **w** counts as similar to the actual world need to be clarified. Such similarity, according to Lewis, does not require that **w** is governed by the same laws of nature as the actual world. The reason is that if world similarity would require sameness of laws, α not occurring in **w** would presuppose that the causes of α are also absent from **w**, as well as their causes, and so forth. In that case, thus, a multitude of states of affairs in **w** would be different from the actual world, meaning that the latter would be very dissimilar from **w**. What is worse, not only effects would counterfactually depend on their causes, but also causes on their effects, for if α is a cause of β, the most similar non-β-world with the same laws would also be a non-α-world. To avoid these consequences, Lewis stipulates that a non-α-world **w** counts as similar to the actual world if all the states of affairs in **w** coincide with the actual world up until the occurrence of α, at which moment a law of nature of the actual world is broken in **w** by a so-called *divergence miracle* suppressing α in **w**. If, and only if, after that miracle, β does not occur in **w**, α and β are counterfactually dependent.

Counterfactual dependence is sufficient but not necessary for causation because counterfactual dependence is not transitive. To ensure that causation is transitive, Lewis (1973) defines causation to be the transitive closure of counterfactual dependence:

(C) α is a token-level cause of β, where α and β are events in the actual world, iff there exists a sequence of events $\langle \alpha, \sigma_1, \ldots, \sigma_n, \beta \rangle$, with $n \geq 0$, such that each event in the sequence counterfactually depends on its predecessor.

As indicated above, (C) is the object of ongoing refinement efforts, as it is subject to various types of counterexamples, for instance, cases of overdetermination or preemption. To illustrate overdetermination, consider this case:

Case 3. *President X issues an executive order. Two judges, independently of one another, overrule that order, which subsequently is suspended. (Each overruling is individually sufficient for the suspension.)*

It seems (pre-theoretically) clear that the rulings of both judges cause the suspension of the order. However, there is no sequence of counterfactual dependencies from either of the rulings to the order's suspension: had the first judge not overruled, the order would still have been suspended, due to the second judge's ruling, and vice versa. That is, (C) erroneously entails that the suspension of the order is caused by neither overruling.

While recent modifications of (C) can adequately handle many of these types of counterexamples, the more fundamental problem remains that possible worlds are not epistemically accessible to us. Determining what would have happened, had certain events been suppressed by a miracle is, to a large degree, a matter of speculation. On the face of it, anything might happen in worlds that feature miracles.

Interventionist theory

The so-called *interventionist* framework provides another non-reductionist approach to defining causation. Interventionist theories have a lot of intuitive appeal and enjoy growing popularity. They tie the notion of causation directly to the way we commonly discover causation: by suitably manipulating factors of interest. Causes are those factors that can be manipulated in such a way that other factors, the effects, vary as well. While some theories in that framework anchor causation in human agency (e.g. Menzies and Price 1993) and, as a result, are subject to anthropocentricity worries, the most compelling interventionist theories define causation in terms of a technical notion of an intervention that is not limited to human agency (e.g. Woodward 2003; Pearl 2009). This section presents the currently most widely used interventionist theory, which is due to Woodward (2003).

Woodward chooses general causation as primary analysandum, develops an anti-realist difference-making theory and neither assumes causation to be deterministic nor transitive. He also remains non-committal with respect to the ontology of causation by referring to causes and effects as *variables* or *factors*—in contrast to, say, regularity theories that take causes and effects to be factor *values*. The difference between factors and factor values seems small but is substantial. A factor A is causally relevant for a factor B (according to Woodward) iff at least one of A's values can make a difference to at least one of B's values, but in order for A=1 to be causally relevant to B=1, not any value of A but the specific value 1 must be the difference-maker for the specific value 1 of B. Put differently, a theory analyzing causation between factors applies to cases such as 'the electoral system is causally relevant for women's representation in parliament', while a theory opting for factor values applies to 'a PR electoral system is causally relevant for high women's representation in parliament'. Hence, in what follows I no longer use the Boolean shorthand notation according to which upper case letters A, B, etc. stand for factor values; rather, they now stand for variables or factors (simpliciter).

The basic idea behind Woodward's interventionist theory is to first specify an ideal test setup $\mathcal{T}$, such that interventions occurring in $\mathcal{T}$ can recover all and only the causal relationships, and to then define A to be causally relevant to B iff that causal relation can be recovered in a $\mathcal{T}$-test. The crucial requirement $\mathcal{T}$ has to comply with in order to serve its designated purpose is *non-confounding*. That is, if the causal relevance of A for B is investigated in $\mathcal{T}$, it must be ensured that no uncontrolled causes of B other than the ones on a path from A to B, *viz.* no *off-path* causes of B, are operative in the background. If A is intervened on while some off-path cause produces B, the resulting data is confounded to the effect that A erroneously appears to make a difference to B. Hence, all off-path causes of scrutinized effects must be homogenized, that is, held fixed in a $\mathcal{T}$-test.

Furthermore, rigorous constraints must be imposed on the way A is intervened on. If the manipulation of A also influences B on a separate path (such that A and B are parallel effects of that manipulation), the resulting co-variation of A and B is confounded and cannot be taken as evidence for the causal relevance of A for B. Likewise, the manipulation of A must not itself be the effect of an off-path cause of B, for that would, again, introduce confounding. In that light, Woodward defines the following technical notion of an *intervention* (Woodward 2003, 98): an intervention on A with respect to B is a surgical cause of A—meaning A is its only direct effect—that is independent of all off-path causes of B.

Against that background, Woodward's (2003) interventionist theory then amounts to the following substitution instance of schema (Φ):

(I) A is a type-level cause of B iff there exists a possible intervention on A with respect to B that is associated with a change in B when all off-path causes of B are held fixed.

Some features of that theory deserve separate emphasis. First, it is non-reductive because it defines causation in terms of interventions, which are certain designated *causes*. According to (I), determining whether A is cause of B requires some prior causal knowledge about the causes of A and B, not however about the relationship between A and B itself (i.e. the theory is not circular). That means (I) can only be brought to bear when that required causal knowledge is available. An interventionist theory cannot analyze causal phenomena from scratch.

Second, even though the notion of an intervention is commonly associated with human action, Woodward's technical definition of the term carries no such connotations whatsoever. Any surgical cause of A, whether the result of human action in an experiment or occurring beyond the range of human action, can count as an intervention. That means, in particular, that an interventionist theory can be applied both in experimental research contexts and in purely observational ones. If observational data contain information about a surgical cause of some test factor A, (I) can be brought to bear.

Third, for A to count as a cause of B, (I) does not require that A be shown to co-vary with B as a result of an *actual* intervention, rather, it suffices that there exists a *possible* intervention establishing A's relevance for B. Whenever

no intervention on A actually occurs, the question thus arises whether such an intervention would be possible and what would happen to B were such an intervention to occur. Answering those questions calls for recourse to possible world semantics, which, in turn, has led to interventionist theories being characterized as (type-level) variants of counterfactual theories.

The main problem of (I) stems from the fact that it not only treats recoverability by an interventionist test as sufficient for causation—which it uncontroversially is—but also as necessary for it. (I) entails that, if it is impossible to surgically cause A (i.e. to intervene on A), A is causally inert. Given the enormous causal complexity of the world we live in, it might well be quite common for factors to be interconnected with so many other factors that they cannot be caused surgically. That such factors do not cause anything does not seem adequate.

Mechanistic theory

Finally, contrary to all previously discussed theories, *mechanistic* theories subscribe to causal realism, that is, to the thesis that not only the causes and effects exist but also the relation connecting them. There are two types of mechanistic theories: *process* theories (e.g. Dowe 2000) and *complex systems* theories (e.g. Glennan 1996). Process theories define causation in terms of the transmission of energy or momentum from causes to effects. They are tailor-made to account for causation in physical systems and are only of minor relevance for social and political science. Accordingly, this section focuses on complex systems theories, which contend that causes are connected to their effects via a mechanism—a complex system of interacting parts—accounting for the production of the effect by the cause. According to this approach, the characteristic feature of causes is not that they are difference-makers of their effects but that they produce them. As production is a causally loaded notion, resulting theories are non-reductive. Mechanistic theories typically remain non-committal as to whether causation is transitive or deterministic. By contrast, they take a decisive stance on the ontology of causation: causes and effects are spatiotemporally located entities, that is, events. Correspondingly, their primary analysandum is singular causation.

The conceptual core of a mechanistic theory is the notion of a mechanism. As indicated above, complex systems approaches do not interpret that notion in a narrow physical sense. Rather, any complex system featuring suitably arranged and interacting parts can count as a mechanism. The following is the most frequently cited characterization of a mechanism in the wide sense; it is due to Machamer et al. (2000, 3):

> Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions.

The deliberately unspecific terms ‘entities’ and ‘activities’ referring to the constituents of a mechanism can be specified for various fields of application. For

instance, Little (2011, 273) stipulates that social mechanisms are 'constituted by the purposive actions of agents within constraints'.

Against that background, Glennan (1996) puts forward the following definition of singular causation:

(M) α is a token-level cause of β, where α and β are non-fundamental events, iff there exists a mechanism connecting α and β.

In other words, α and β are causally related iff there is a sequence of intermediary events, each of which is caused (produced) by its predecessor and causes (produces) its successor. The causal relations among the elements of that sequence are again to be analyzed in terms of the existence of intermediary mechanisms, and so forth. That is, (M) non-reductively spells out causation between upper-level events in terms of causation between lower-level events, giving rise to a regress that continues until a fundamental level is reached where nothing intermediary exists any longer, if such a level exists; or, if no fundamental level exists, the regress continues ad infinitum. In any case, causation between events on a fundamental level cannot be understood in the vein of (M), which is why Glennan restricts the applicability of (M) to non-fundamental events. Causation between fundamental events must then either be analyzed on the basis of another theory of causation or treated as pre-theoretically clear (i.e. fundamental).

As all other non-reductive theories, (M) can only be applied if a significant amount of prior causal knowledge is already available. While other non-reductive theories, however, presuppose clarity on causation in the background (i.e. outside) of a scrutinized causal relation, (M) presuppose knowledge about what is going on in between candidate causes and effects (i.e. inside of a causal relation).

5 Methodological perspectives

I end this article by briefly connecting the above theories to available methods of causal inference and discovery. As documented in this part of the Handbook, there exists a multitude of methods. Many of them are tailored to uncover causation as defined by different theories. Given the vast amount of available methodological frameworks, I cannot render explicit for all of them what type of causation they trace. Hence, without claim to completeness, the following paragraph links every theory discussed in this article to exemplary methods designed to uncover causation as defined by that theory.

Regularity theoretic causation can be uncovered by configurational comparative methods (e.g. Ragin 2008, Baumgartner and Ambühl 2018). Bayesian network methods (e.g. Spirtes et al. 2000) and regression analytic methods (e.g. Gelman and Hill 2007) trace the probabilistic variant of causation. The potential outcomes framework and structural equation modeling (e.g. Morgan and Winship 2007, Halpern 2016) provide useful tools to search for causation in the sense of counterfactual theories. Interventionist causation can be uncovered by experimental methods or by interventionist variants of Bayesian network methods or

structural equation modeling (e.g. Pearl 2009). And a paradigmatic framework to examine mechanistic causation is process tracing (e.g. Beach and Pedersen 2013).

According to the position of causal pluralism, the theories of causation discussed in this article define different concepts of causation by doing justice to divergent (and often incompatible) properties pre-theoretically ascribed to causation. It follows that there is no fact of the matter as to which is the true theory. Causal pluralism yields methodological pluralism. Methods tracking causation as defined by different theories have different search targets and, consequently, cannot be meaningfully pitted against each other. There is no fact of the matter which of them is more truth-conducive, rather, they complement one another. The same does not hold for methods targeting the same variant of causation. Such methods can and must be rigorously benchmarked against each other. If one of them turns out to more correctly and completely uncover the relevant variant of causation it is strictly superior.

The choice of theory and the choice of method mutually constrain each other and are constrained by the exigencies of a given research context. If causally analyzed data are of correlational nature, a method is called for that can process correlational data and causation must be understood in terms of a theory that connects correlational dependencies to causation. Appropriate choices in such a context are a probabilistic theory and a method from the Bayesian network or regression analytic framework. If a study is examining the complexity of the type-level causal structure underlying a phenomenon of interest, a theory and corresponding method are needed that are capable of reproducing conjunctural causation, equifinality, and causal sequentiality. In that case, a regularity theory might be chosen along with a pertinent configurational comparative method. Alternatively, if enough prior knowledge on how to intervene on an investigated phenomenon is available, the interventionist framework is a safe choice. Or if a study aims to explain the occurrence of some token-level effect, suitable choices might be a counterfactual theory in combination with the potential outcomes framework or process tracing underwritten by a mechanistic theory.

References

Albert, D. Z. (1992). *Quantum Mechanics and Experience*. Cambridge: Harvard University Press.

Baumgartner, M. (2013). A regularity theoretic approach to actual causation. *Erkenntnis 78*, 85–109.

Baumgartner, M. and M. Ambühl (2018). Causal modeling with multi-value and fuzzy-set Coincidence Analysis. *Political Science Research and Methods*. doi: 10.1017/psrm.2018.45.

Baumgartner, M. and C. Falk (2018). Boolean difference-making: a modern regularity theory of causation. *PhilSci Archive*. url: http://philsci-archive.pitt.edu/id/eprint/14876.

Beach, D. and R. Pedersen (2013). *Process tracing methods: foundation and guidelines*. University of Michigan Press.

Beebee, H., C. Hitchcock, and P. Menzies (Eds.) (2009). *The Oxford Handbook of Causation*. Oxford: Oxford University Press.

Bunge, M. (1979). *Causality and modern science* (Third revised edition ed.). New York: Dover Publications.

Cartwright, N. (1979). Causal laws and effective strategies. *Noûs 13*, 419–437.

Casati, R. and A. Varzi (2015). Events. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2015 ed.). Metaphysics Research Lab, Stanford University.

Davidson, D. (1967). Causal relations. *Journal of Philosophy 64*(21), 691–703.

Dowe, P. (2000). *Physical Causation*. Cambridge: Cambridge University Press.

Eells, E. (1991). *Probabilistic Causality*. Cambridge: Cambridge University Press.

Ehring, D. (2010). Causal relata. In H. Beebee, C. Hitchcock, and P. Menzies (Eds.), *The Oxford Handbook of Causation*, pp. 387–413. Oxford University Press.

Gelman, A. and J. Hill (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis 44*, 49–71.

Glynn, L. (2011). A probabilistic analysis of causation. *British Journal for the Philosophy of Science 62*, 343–392.

Graßhoff, G. and M. May (2001). Causal regularities. In W. Spohn, M. Ledwig, and M. Esfeld (Eds.), *Current Issues in Causation*, pp. 85–114. Paderborn: Mentis.

Halpern, J. (2016). *Actual Causality*. Cambridge, MA: MIT Press.

Halpern, J. Y. and C. Hitchcock (2010). Actual causation and the art of modelling. In R. Dechter, H. Geffner, and J. Y. Halpern (Eds.), *Heuristics, Probability, and Causality*, pp. 383–406. London: College Publications.

Healey, R. (2010). Causation in quantum mechanics. In H. Beebee, C. Hitchcock, and P. Menzies (Eds.), *The Oxford Handbook of Causation*. Oxford University Press. doi: 10.1093/oxfordhb/9780199279739.003.0034.

Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy 98*, 273–299.

Hume, D. (1999 (1748)). *An Enquiry Concerning Human Understanding*. Oxford: Oxford University Press.

Lewis, D. (1973). Causation. *Journal of Philosophy 70*, 556–567.

Lewis, D. (2000). Causation as influence. *Journal of Philosophy 97*, 182–197.

Little, D. (2011). Causal mechanisms in the social realm. In P. M. Illari, F. Russo, and J. Williamson (Eds.), *Causality in the Sciences*, pp. 273. Oxford University Press.

Machamer, P. K., L. Darden, and C. F. Craver (2000). Thinking about mechanisms. *Philosophy of Science 67*(1), 1–25.

Mackie, J. L. (1974). *The Cement of the Universe. A Study of Causation*. Oxford: Clarendon Press.

Mellor, D. H. (1995). *The Facts of Causation*. London: Routledge.

Menzies, P. and H. Price (1993). Causation as a secondary quality. *British Journal for the Philosophy of Science 44*, 187–203.

Morgan, S. and C. Winship (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Analytical Methods for Social Research. Cambridge: Cambridge University Press.

Pearl, J. (2009). *Causality. Models, Reasoning, and Inference* (2 ed.). Cambridge: Cambridge University Press.

Psillos, S. (2010). Causal pluralism. In R. Vanderbeeken and B. D'Hooghe (Eds.), *Worldviews, Science and Us: Studies of Analytical Metaphysics*, pp. 131–151. World Scientific Publishers.

Ragin, C. C. (2008). *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: University of Chicago Press.

Reichenbach, H. (1956). *The Direction of Time*. Berkeley: University of California Press.

Russo, F. and J. Williamson (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science 21*(2), 157–170.

Spirtes, P., C. Glymour, and R. Scheines (2000). *Causation, Prediction, and Search* (2 ed.). Cambridge: MIT Press.

Suppes, P. (1970). *A Probabilistic Theory of Causality*. Amsterdam: North Holland.

Tooley, M. (1987). *Causation: A Realist Approach*. Oxford: Clarendon Press.

Woodward, J. (2003). *Making Things Happen. A Theory of Causal Explanation*. New York: Oxford University Press.