Lars-Göran Johansson · Thomas Banitz ·
Volker Grimm · Tilman Hertz · Emilie Lindkvist ·
Rodrigo Martínez Peña · Sonja Radosavljevic ·
Petri Ylikoski · Maja Schlüter

A Primer to Causal Reasoning About a Complex World

Springer

SpringerBriefs in Philosophy

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic. Typical topics might include:

- A timely report of state-of-the art analytical techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- An in-depth case study or clinical example
- A presentation of core concepts that students must understand in order to make independent contributions

SpringerBriefs in Philosophy cover a broad range of philosophical fields including: Philosophy of Science, Logic, Non-Western Thinking and Western Philosophy. We also consider biographies, full or partial, of key thinkers and pioneers.

SpringerBriefs are characterized by fast, global electronic dissemination, standard publishing contracts, standardized manuscript preparation and formatting guidelines, and expedited production schedules. Both solicited and unsolicited manuscripts are considered for publication in the SpringerBriefs in Philosophy series. Potential authors are warmly invited to complete and submit the Briefs Author Proposal form. All projects will be submitted to editorial review by external advisors.

SpringerBriefs are characterized by expedited production schedules with the aim for publication 8 to 12 weeks after acceptance and fast, global electronic dissemination through our online platform SpringerLink. The standard concise author contracts guarantee that

- an individual ISBN is assigned to each manuscript
- each manuscript is copyrighted in the name of the author
- the author retains the right to post the pre-publication version on his/her website or that of his/her institution.

Lars-Göran Johansson • Thomas Banitz •
Volker Grimm • Tilman Hertz • Emilie Lindkvist •
Rodrigo Martínez Peña • Sonja Radosavljevic •
Petri Ylikoski • Maja Schlüter

A Primer to Causal Reasoning About a Complex World

Springer

Lars-Göran Johansson
Department of Philosophy
Uppsala University
Uppsala, Sweden

Volker Grimm
Department of Ecological Modelling
Helmholtz Centre for Environmental
Research - UFZ
Leipzig, Germany

Emilie Lindkvist
Stockholm Resilience Centre
Stockholm University
Stockholm, Sweden

Sonja Radosavljevic
Stockholm Resilience Centre
Stockholm University
Stockholm, Sweden

Maja Schlüter
Stockholm Resilence Centre
Stockholm University
Stockholm, Sweden

Thomas Banitz
Department of Ecological Modelling
Helmholtz Centre for Environmental
Research - UFZ
Leipzig, Germany

Tilman Hertz
Stockholm Resilience Centre
Stockholm University
Stockholm, Sweden

Rodrigo Martínez Peña
Institute of Analytical Sociology
Linköping University
Norrköping, Sweden

Petri Ylikoski
Department of Sociology
University of Helsinki
Helsinki, Finland

The development of Western science is based on two great achievements; the invention of the formal logical system (in Euclidian geometry) by the Greek philosophers, and the discovery of the possibility to find out causal relationships by systematic experiments (during the Renaissance).
A. Einstein (1953)

No causes in, no causes out.
N. Cartwright (1989)

Preface

This book is meant as an introduction to students and researchers in the multi-disciplinary field of sustainability science who are interested in a better understanding of what 'causation' means, and how it can be explored in a more systematic way. It introduces fundamental ideas about causation from philosophy, particularly those that underlie studies of causation that are based on quantitative data and statistical methods of causal inference. Chapter 9 then takes a broader view to present the diversity of causal reasoning found in Social Ecological Systems, SES for short research and discuss how to disclose and navigate it. While we are focusing on application to, SES, where such better understanding is badly needed, readers from virtually all disciplines will benefit, because in most disciplines assumptions about causation are usually implicit, which often will limit progress. Thus, this book should serve as an introductory textbook to be used for classes and seminars, but also for self-studying. To support this, each chapter ends with a list of questions that guide to further literature, stimulate discussions, or to be used for homework assignments.

This book is one result of an interdisciplinary project at Stockholm Resilience Centre at Stockholm University, 'Approaches to causation in the social and natural sciences and their implications for theory building in sustainability science—CauSES', whose goal was to investigate causal thinking in research on SES. This is a relatively new interdisciplinary field, attracting researchers with different backgrounds and different presuppositions about causal relations. Observing the vast differences between ecologists, physicists, economists, political scientists, sociologists, anthropologists, statisticians and others regarding causation, one might wonder whether there is any common idea about causation at all.

Philosophers have discussed both epistemological and a metaphysical aspects of causation since antiquity. The debate has been more lively than ever the last decades. There is some convergence about some aspects of causation, but universal agreement about the concept of causation is still not in sight.

Our approach is more empirical than most philosophical contributions to the causation debate. We have two starting points: the first is exploring how we use the terms 'cause', 'effect' and related words in ordinary discourse without any explicit

justification. These uses together make up an implicit characterisation of the basic meaning of causal terms. The second is investigating how scientists in different disciplines make inferences using the causal vocabulary.

A fairly coherent conception of causal relations has been achieved in the natural sciences. The extension of this conception to parts of social sciences has been difficult, for reasons to be discussed in this book.

We, the nine members of the research group, have had many meetings where outlines of papers and chapters of this book have been thoroughly discussed. It was decided that Lars-Göran Johansson should take the main responsibility for writing this book by writing drafts of the chapters, which then have been discussed several rounds in the entire group. Chapter 9 is an exception; here Maja Schlüter, Emilie Lindkvist, Tilman Hertz, Rodrigo Martínez Peña and Thomas Banitz wrote the first drafts.

Uppsala, Sweden	Lars-Göran Johansson
Leipzig, Germany	Thomas Banitz
Leipzig, Germany	Volker Grimm
Stockholm, Sweden	Tilman Hertz
Stockholm, Sweden	Emilie Lindkvist
Norrköping, Sweden	Rodrigo Martínez Peña
Stockholm, Sweden	Sonja Radosavljevic
Helsinki, Finland	Petri Ylikoski
Stockholm, Sweden	Maja Schlüter

Contents

Chapter 1
Introduction: Causation in Social-Ecological Systems

Abstract In this chapter we start the discussion about causal idiom by giving excerpts from three papers, each discussing the dynamics of a social-ecological system. There is plenty of talk about causes in these papers, but, interestingly, the authors talk about causes and effects without much reflection on the criteria for something being a cause of something else, nor about the required evidence for such claims.

This book is about causal thinking and use of causal idiom in general, with the aim to provide the basic understanding required to explore the diversity of causal reasoning about social-ecological systems (SES) in particular.[1] Three questions stand in focus: (i) What are the meanings of different causal expressions, (ii) what is sufficient evidence for inferences from observations to causal relations, and (iii) how to handle the diversity of causal relations in SES?

As a starter, let's have a brief look at three excerpts from SES research. The first is from (Hruska, 2017):

> A social–ecological system (SES) is a combination of social and ecological actors and processes that influence each other in profound ways. The SES framework is not a research methodology or a checklist to identify problems. It is a conceptual framework designed to keep both the social and ecological components of a system in focus so that the interactions between them can be scrutinised for drivers of change and causes of specific outcomes. Resilience, adaptability, and transformability have been identified as the three related attributes of SESs that determine their future trajectories. Identifying feedbacks between social and ecological components of the system at multiple scales is a key to SES-based analysis.
>
> ...[T]he SES framework facilitates identification of cross-system feedbacks to explain otherwise puzzling outcomes. While information intensive and logistically challenging in the management context, the SES framework can help overcome intractable challenges to working rangelands such as rangeland conversion and climate change. The primary benefit

---

[1] Illari and Russo (2014) and Norton et al. (2014) are two other books discussing scientific practice and philosophical theories about causation.

of the SES framework is the improved ability to prevent or correct social policies that cause negative ecological outcomes, and to achieve ecological objectives in ways that support, rather than hurt, rangeland users. (op. cit. p. 263)

Several expressions indicating causal relations are here used: 'interactions', 'drivers of change', 'causes', 'feedbacks', 'prevent', 'correct' and 'achieve objectives'. The authors do not give any precise definitions of these expressions, apparently thinking that they are sufficiently common and well understood in ordinary language use or within the SES community.

Later in the chapter we find a diagram showing the components of rangeland social-ecological systems, but it is no diagram in the ordinary sense of this word; It is no more than a display of a number of concepts ordered in five or six groups. It does not suggest anything about causal relations.

The caption says a lot more, but almost nothing of what is said in that text is displayed in any way in the figure (Fig. 1.1).

Fig. 1.1 Generalized diagram of a rangeland social–ecological system. Humans and the environment interact in countless ways outside of natural resource management, but the interactions are most directly planned, manipulated, and monitored in natural resource management activities. Local, regional, and global social processes can all shape natural resource use and management activities. While resource policy may be set at large geographical scales (e.g., national), management activities occur within a single ecosystem. Livestock grazing differs from other types of natural resource use in that it is indirect; rather than directly manipulating a rangeland ecosystem, livestock operators devote their primary attention to managing livestock, and the livestock interact directly with the rangeland (adapted from Hruska, 2017, 266)

The second excerpt is a case study of green turtle fisheries in Nicaragua. Here is the summary:

2.3 Robustness Summary. The Nicaraguan green turtle fishery does not represent a robust system of CPR governance. Persistent poverty, lack of alternative employment opportunities, and a high population growth rate (initially within communities, but more recently through in-migration) continue to be the main drivers for the commercialization of this fishery under a domestic subsistence use exception to endangered species protection of green turtles. Instead of protecting the species from exploitation, the ratification of the Convention on International Trade in Endangered Species (CITES) in 1977 by the national government, and subsequent closure of the legal international market for green turtle meat, has merely led to a shift in focus by turtle fishers from responding to international market demands to creating and satisfying a domestic demand for green turtles. Although de jure rules limiting the harvesting of green turtles exist at all government levels, including at the territorial, municipal, and community level in the RAAS and the RAAN, there is no overarching coordination of those rules, and no monitoring or enforcement, but for the collection of harvesting data that is being conducted by a researcher formerly involved with an international NGO. In essence, the fishery is de facto operated year-round without any restrictions. Prior limiting factors, such as the special skills required to navigate sailing dories and harpoon turtles, have been eliminated through the increased use of motor boats and turtle nets. The literature mentions three factors that provide evidence of the long-term unsustainability of the fishery: (1) actual capture rates are believed to be significantly higher than reported; (2) a majority of the animals captured are large, sexually immature juveniles and adult turtles from the Tortuguero natal nesting site, which effectively removes the base for a future breeding population; and (3) recent declines in capture rates in regions with previous turtle abundance (Lagueux et al. 2014). Given many Miskitos reliance on green turtles as a sole source of cash revenues, a turtle population collapse could have significant social-ecological consequences (Brady et al., 2015).

The core causal claim is that there are 'three factors that provide evidence of the long-term unsustainability of the fishery'. This formulation tells us that plausibly there are three *causes* of unsustainability, although the word 'cause' is not used. It is followed by a short account of the causal mechanisms.

The third excerpt is the abstract of a paper about the governance of coastal fisheries in Chile:

Here we explore social, political, and ecological aspects of a transformation in governance of Chile's coastal marine resources, from 1980 to today. Critical elements in the initial preparatory phase of the transformation were (i) recognition of the depletion of resource stocks, (ii) scientific knowledge on the ecology and resilience of targeted species and their role in ecosystem dynamics, and (iii) demonstration-scale experimental trials, building on smaller-scale scientific experiments, which identified new management pathways. The trials improved cooperation among scientists and fishers, integrating knowledge and establishing trust. Political turbulence and resource stock collapse provided a window of opportunity that triggered the transformation, supported by new enabling legislation. Essential elements to navigate this transformation were the ability to network knowledge from the local level to influence the decision-making processes at the national level, and a preexisting social network of fishers that provided political leverage through a national confederation of artisanal fishing collectives. The resultant governance scheme includes a revolutionary national system of marine tenure that allocates user rights and responsibilities to fisher collectives. Although fine tuning is necessary to build resilience of this new regime, this transformation has improved the sustainability of the interconnected social–ecological system. Our analysis of how this transformation unfolded provides insights into how the

Chilean system could be further developed and identifies generalised pathways for improved governance of marine resources around the world (Gelcich et al., 2010).

In all three examples it is clear that a main goal of the research is to arrive at knowledge about system's dynamics and what to do to change things, in the first case to improve management of rangelands, in the second case to arrive at a more sustainable turtle fishery and in the third case to explain the change in governance of coastal fisheries. In order to know what to do one needs causal knowledge, reliable knowledge of the form, 'If we do X, then Y will probably occur'.

Causal knowledge is thus often wanted because one wants to understand why and how SES change and to obtain guidance for future actions. However, in neither case is the reported research particularly illuminating about the criteria for counting something as a cause of something else. And with one exception, there is no indication of what kind of evidence the authors require for inferences to causes of observed states of affairs.

This, we believe, is a rather common feature of much empirical science; causal notions are often used without much reflection and questions about evidence for causal relations are often not discussed. The intense discussion among philosophers about the concept of causation has had little influence on empirical scientists. This we will try to remedy to some extent.

Broadly speaking there are three types of questions addressed by SES researchers:

1. What are the causes of a particular undesirable state of affairs?
2. What are the effects of a possible intervention in a social-ecological system?
3. What is the causal structure of a SES?

Answers to questions 1 and 2 are causal explanations, while answers to the third one are *constitutive explanations*, i.e., explanations that consist of descriptions of the parts making up the system. This distinction will be elaborated upon in Chap. 8.

The rest of this book is divided into three parts. The Part I, Chaps. 2, 3, and 4, is a general analysis and discussion of the use of causal idiom in ordinary language. It does not result in an explicit definition of the terms 'cause' and 'effect', but it gives their meaning in particular contexts. This is a necessary prerequisite for a discussion about causal inference in science, which is the topic of the Part II, Chaps. 5–8. The Part III, Chap. 9, is about the diversity of causal reasoning in SES research. It discusses how this diversity results from the diverse backgrounds, interests, epistemological stances and scientific norms of researchers and practitioners in this multi-disciplinary field and where and how during a research process it manifests. It also provides suggestions about how to navigate this often confusing diversity.

While the chapters do have, of course, a sequence, and although we suggest to follow this sequence while reading, there is not necessarily a strict linear flow from simple to complex, or basic to advance, as a textbook on, e.g., statistics would be organised. Rather, there are multiple aspects to consider, when trying to understand what 'causation' can mean, and what the challenges are when trying to understand causation. Therefore, to some degree, each chapter has its own main focus, which

is relevant but also strongly connected to most of the topics of the other chapters. To make this very clear, we added, after each chapter's abstract, a bullet point list summarising the main lessons to be learned in this chapter, plus a brief summary why and how the chapter's topics are relevant regarding the overall purpose of this book, which is being better prepared to ask and answer questions about causation, both in general and for SES in particular.

References

Brady U, Tellman B, Anderies JM (2015) Institutional analysis of the artisanal green turtle fishery in eastern (Caribbean) Nicaragua

Gelcich S, Hughes TP, Olsson P, Folke C, Defeo O, Fernández M, Foale S, Gunderson LH, Rodríguez-Sickert C, Scheffer M, Steneck RS, Castilla JC (2010) Navigating transformations in governance of chilean marine coastal resources. Proc Natl Acad Sci 107(39):16794–16799

Hruska T et al (2017) Rangelands as social–ecological systems. In: Briske D (ed) Rangeland systems. Processes, management and challenges, chap 8. Springer, Berlin

Illari P, Russo F (2014) Causality: philosophical theory meets scientific practice. Oxford University Press, Oxford

Norton SB, Cormier SM, Suter GW II (2014) Ecological causal assessment. CRC Press, Boca Raton

Part I
Semantics of Causal Expressions

Chapter 2
Causal Talk Permeates Ordinary Language

Abstract This chapter gives an overview of causal idiom in ordinary language and introduces some fundamental semantic distinctions. The main points are:

- The words 'cause' and 'effect' have quite a number of near synonyms.
- 'Cause' and 'effect' are relational terms.
- Causal relations relate three kinds of things: events, categories of events and variables used in statistical treatments of such categories.
- A mathematical relation between variables does not in itself tell us whether they are causally related or not.

2.1   Introduction

There is a great variety of causal expressions and even a very brief overview of ordinary language reveals how common expressions for causal relations are; some examples are 'bring about', 'make happen', 'produce', 'do', 'perform' 'result in' 'effect of', and 'leads to'. Two questions immediately come to the fore: (i) is there any common meaning of these expressions, a core meaning of respectively 'cause' and 'effect', and (ii) what evidence is required as justification for causal claims?

Many different things have been related as cause and effect in ordinary as well as in philosophical and scientific discourse: events, states of affairs, properties, categories, quantities, processes, desires, beliefs and actions. This list is not complete and one may reasonably wonder if there is any core meaning at all for all the diverse uses of the terms 'cause', effect' and other terms with a causal connotation. We believe there is, there are some necessary conditions for the correctness of sentences of the form 'x causes y', and its converse 'y is the effect of x', but in any particular context more conditions are needed and these differ in different contexts. There is a wide diversity of causal notions, which will be surveyed and analysed in this book.

Causal idiom and causal thinking is a basic trait of humans. Already in the second verse of Iliad (circa 800 BC), the first work of western literature, We are given a seemingly causal explanation of the hostility between Agamemnon and Achilles:

> And which of the gods was it that set them on to quarrel?
> It was the son of Jove and Leto; for he was angry with the king
> and sent a pestilence upon the host to plague the people,
> because the son of Atreus had dishonoured Chryses his priest. (Transl. S. Butler).

Thus, the conflict between Agamemnon and Achilles is said to have been caused, indirectly, by the god Apollo ('the son of Jove an Leto') who is attributed a typical human psychology: Apollo was angry with the king and acted accordingly. The cause of the quarrel is an action of a sentient being.

In the epilogue to his book *Causality: Models, Reasoning and Inference* Judea Pearl does not mention this passage, but noticed that in ancient times questions about causes were questions about agents, their motives and desires:

> The agents of the causal forces in the ancient world were either deities, who cause things to happen for a purpose, or human beings and animals, who possess free will for which they are punished and rewarded. This notion of causation was naive but clear and unproblematic.
>
> The problem began, as usual, with engineering; when machines had to be constructed to do useful jobs. [....] And once people began to build multistage systems, an interesting thing happened to causality - *physical objects began acquiring causal character.* (Pearl, 2000, 333)

One might say that in ancient times the prototypical cause is an agent who acts for a purpose, whereas from the scientific revolution onwards the prototypical case is a ball colliding with another ball and changing the latter's motion.

The ancient notion of causation—a cause is an action of an agent—is still very common in ordinary thinking and language, less so in scientific discourse, except perhaps in some social sciences. The latter are concerned with human actions, individual and collective, so in these disciplines there is plenty of talk about causal agents driven by beliefs and desires.

Having analysed the different aspects of causal discourse in both ordinary and several scientific contexts, we will at the end of the book focus on causal reasoning in complex SES. Analysing causal relations in such systems is particularly demanding because of the diversity of kinds of entities, properties and relations and the complex dynamics they generate.[1]

---

[1] A book that focuses on our habits and cognitive abilities to understand causal connections is (Grotzer, 2012).

2.2   Causal Phrases

As an illustration of the variety of expressions used for making causal claims, we may have a look at the beginning of (Lindegren et al., 2009):

> Atlantic cod (Gadus morhua) is among the commercially most important fish species of the European waters. Many of the stocks have declined dramatically and still remain at historically low levels (1, 2). These collapses have largely resulted from overfishing (3, 4) and climate-driven declines in productivity (5, 6). The climate effect generally works through changes in the physical environment (e.g., temperature and salinity), but also through altered food supply for early life-history stages, eventually affecting recruitment (5, 6). In accordance with this effect, recruitment failure of Eastern Baltic cod was caused mainly by high egg and larval mortalities as a result of climate-induced hydrographic change (7, 8). In several areas the collapses of cod stocks were part or major drivers of large-scale reorganisations of ecosystems (9). These so-called regime shifts are frequently caused by climatic changes (9, 10) and/or over-exploitation resulting in cascading trophic interactions (11, 12). Similarly to other areas, the Baltic Sea underwent both regime shifts and trophic cascades (8, 13). Such alterations in ecosystem structure typically affect species interactions, eventually influencing food-web dynamics through both positive and negative feedback loops (14).

In this introduction we observe at least seven expressions indicating causal relations:

1. 'These collapses have largely *resulted* from overfishing.'
2. '...climate *driven* declines in productivity...'
3. 'The climate *effect* generally works through changes in the physical environment.'
4. '...failure of Eastern Baltic cod was *caused* mainly by high egg and larval mortalities.'
5. '....the collapses of cod stocks were part or major *drivers* of large-scale reorganisations of ecosystems.'
6. 'These so-called regime shifts are frequently *caused* by climatic changes.'
7. 'Such alterations in ecosystem structure typically *affect* species interactions, eventually *influencing* food-web dynamics through both positive and negative feedback loops.'

All these sentences convey information having the form that someone or something makes something else happen. In neither case could the researchers directly have observed these events, so one immediately wonders how the authors know that the relations expressed by 'cause', 'make happen', 'drives', 'affect', etc, really obtain, i.e. what evidence they have collected. This question in turn triggers the question what these expressions really mean, for we cannot decide what evidence we require for a certain statement if we do not know what it means. Questions about evidence for causal claims and questions about the meaning of these claims are thus deeply and intimately related; the meaning of an expression determines what kinds of evidence there might be for a sentence containing this expression.

The diversity of ideas about causation is not only a matter of methods, there are also different ideas of what causation is and what it entails. For instance, people

disagree about (i) generalisability of singular causal relations, (ii) origins of causal powers, and (iii) which causes are more important than others.

2.3 Some Remarks on the Semantics of 'Cause', 'Effect' and Their Cognates

2.3.1 Causal Relations Between Events/States of Affairs

Our most basic use of causal idiom consists in relating *particular* events or states of affairs by a two-place predicate 'x caused y', 'x was the effect of y', 'x leads to y', or some other expression with clear causal meaning.[2] In other words, things causally related to each other are singular events or states of affairs that occur at particular times and places. This is the fundamental use of causal idiom. But use of words for causal relations is wider: with the development of modern science, causal talk has been extended to cover also relations between categories and quantities.[3] By abstracting from individual cases, we simply say that one attribute is the cause of another attribute. For example, overweight is said to be a cause (not the only one!) of high blood pressure.

2.3.2 Causal Relations Between Categories

The pandemic Covid-19 was caused by the virus SARS-CoV-2. This is a relation between two categories: the cause is a virus of a certain type or category, and the effect is a disease of another category; being exposed to particles belonging to this virus type increases the probability of getting the disease Covid-19.

A necessary condition for this causal relation to occur is that the conditional probability to attract Covid-19 when exposed to the virus is higher than the marginal probability to get the disease. In mathematical notation:

$$prob(|\text{Covid-19} \mid \text{being exposed to SARS-CoV-2}) > prob(\text{Covid-19}). \quad (2.1)$$

---

[2] In logic, the term 'predicate' has a more general meaning than in ordinary grammar. Predicates are what is left of a sentence when you remove the noun phrase and direct objects. If you only remove the noun phrase, you have a one-place predicate, if you remove more terms you get two-place, three-place predicates, etc.

[3] Collected data must be organised in some way. This is done using variables. The basic distinction among variables is between category variables, such as sex, or ethnic group, and quantitative variables such as biomass, weight or age. Hence, categories are basically sets of objects identified by a common attribute, whereas quantities are quantitative attributes of things, see further Sect. 5.2.

This means that Covid-19 and SARS-CoV-2 are correlated. Still, this statistical relation is not sufficient evidence for there being a causal relation; more evidence is needed, see Sect. 2.4 and Chap. 6. In this case we know that the virus is the cause of the infection based on experimental evidence, not just on statistical correlations. More carefully expressed: we know that in any individual case of someone having the Covid-19 infection, this event was caused by him being infected by the SARS-CoV-2 virus. So we generalise and say that the virus causes this disease, thus saying that one category causes another category. We have thus extended the possible relata of the relation ... *causes*.... to include categories. (In this case the disease is identified by its cause, so the probability of getting the disease without being exposed to SARS-CoV-2 is zero.)

From a purely grammatical point of view, the terms 'SARS-CoV-2' and 'Covid-19' are singular terms (see textbox below). That does not conflict with their referents being categories, i.e., classes of things. A category, when given a name, is treated, grammatically and from a logical point of view, as an entity, a particular thing. (Whether this should be understood as that there exists categories, i.e., properties and relations, over and above the individual cases, is a perennial dispute in metaphysics. Luckily, for the purpose of this book, we need not take any stance in this debate.)

Singular and General Terms

From a logical point of view, the simplest cases of complete declarative sentences consist of a general term, (a predicate phrase) and one or several singular terms. Singular terms are proper names, personal pronouns, definite descriptions, or variables standing for such things.

A singular term cannot function as predicate, while general terms can occur both as nouns and predicates.

An example with one singular term is

#1. The oldest person in Sweden is more than 100 years.

Here, 'The oldest person in Sweden' is the singular term, it is a definite description of one distinct entity. The rest of the sentence is the predicate phrase 'is more than 100 years'; this is a general term, which means that it can be predicated about many things.

A sentence with two singular terms is

#2. Sweden has a smaller population than Germany.

Here, 'Sweden' and 'Germany' are the two singular terms, they are each a name of a political and geographical entity. (Thus in predicate logic we do not distinguish between noun and direct object.)

A sentence with three singular terms is

#3. The judge in a civil case determines whether the plaintiff or the defendant wins the case.

Here, 'The judge in a civil case', 'the plaintiff' and 'the defendant' are the three singular terms.

So singular and general terms are the logical constituents of the simplest complete declarative sentences. Singular terms *may* refer to things, objects, events, etc., i.e., entities that can be identified as individuals, i.e., thought of as *one*.

A singular term need not refer to anything. 'The king of France' is a singular term, but there is no king of France.

Things referred to by singular terms, i.e., individuals, need not be individuals in any ordinary sense. In the sentence 'Manchester United won The Premier League 2012–2013', the name 'Manchester United', is a name for an individual entity, a football club. When talking about Manchester United we *treat it as one object*, disregarding whether or not it consists of a number of players and other members.

2.3.3 Causal Relations Between Quantitative Variables

A common scientific question is whether a certain variable is the cause of another variable and huge efforts are often made in order to answer such questions. The starting point when asking about a possible causal relation between two variables is to see whether they are correlated or not. Suppose the answer is yes. That is not sufficient for inferring that they are causally related, since a correlation can occur without there being a causal connection. But if there is a causal relation between two variables, they are correlated, when other variables are controlled for. So observing a correlation is a reason for further inquiry to see whether there is a causal relation or not.

As a starter, we observe that an expression of the form 'variable X is a cause of variable Y' means that a change in the value of variable X attributed to some object causes a change in the value of variable Y attributed to the same or another object. In other words, causal relations between variables are based on causal relations between ordered pairs of individual events. But neither in ordinary nor in scientific contexts is this explicitly stated; the common expression is that a certain variable causes another one.

Relations between two quantitative variables are often expressed as functions of the form $y = f(x)$. Such an expression does not contain any information about any causal relation, because a function by itself does not express any relation between

events in the real world. But even if we add an interpretation to the effect that variable values represent events in the world that are causally related, the form of this expression does not distinguish between cause and effect. The reason is that the causal relation is asymmetric, it is 'directed' from the cause to the effect, while a mathematical function only expresses a numerical relation between the values of the two variables.

In many cases, when the function $y = f(x)$ has an inverse, this is immediately clear, since the equations $y = f(x)$ and $x = f^{-1}(y)$ are logically equivalent, they are two different expressions for the same fact of the matter. Thus, the mere syntactic form of the equation $y = f(x)$ does not tell us anything about which is the cause and which is the effect, or whether there is any causal relation at all between these variables. We will discuss this more thoroughly in the next section. The formalism of *structural equations* is another matter, to be discussed in Chap. 7.

Most often a causal interpretation of an equation comes from an intuitive and tacit judgement about which variable we naturally, or most easily, can manipulate. Hence, the causal interpretation of a mathematical relation between two variables comes from our *agency perspective*, to be discussed in Chap. 3. An equation by itself does not say anything at all about causal relations.

2.3.4   Common Causes

The fact that a correlation between two variables is very strong does not by itself say anything about the probability for there being a causal connection. There are many well-known cases of correlations between two variables that no one would think of as causally related. The correlation is in such a case explained by being produced by a *common cause*, often called a confounder. If variable A is a cause of variable B via one causal mechanism and a cause of another variable C via another causal mechanism, we may observe a correlation between B and C without there being any causal link between them. Here is one example of a strong correlation which most plausibly, given even a very limited background knowledge, is the outcome of a common cause. Tyler Vigen, from whose home page tylervigen.com/spurious/correlation the figure is taken, discusses some possibilities (Fig. 2.1).

The expression 'confounding cause' is often used when referring to a common cause. This term is a bit misleading since it suggests that the confounding cause is not a real cause, while in fact it *is* the cause of the correlation. What the confounder can do is to mislead us into thinking that there is a causal link between two observed and correlated variables, while there is not.

How to decide, by empirical means, whether two observed variables really are causally related or whether there is a common cause will be discussed in Chap. 6.
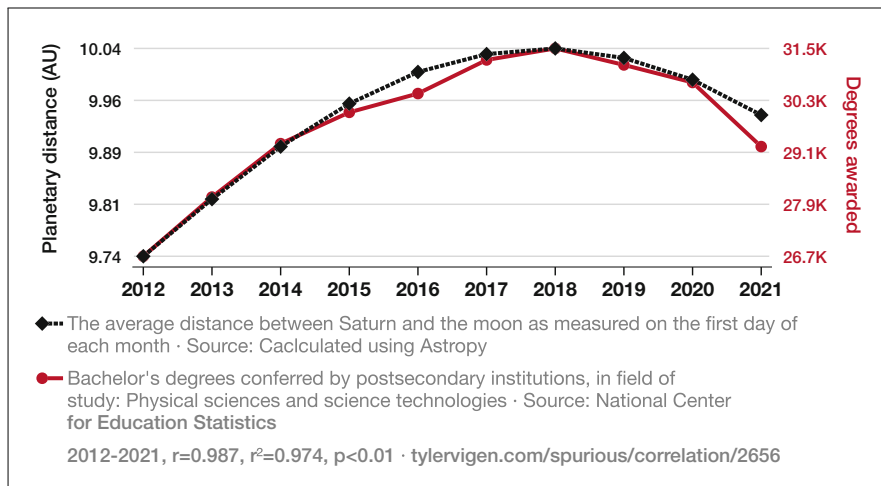
Fig. 2.1 The number of bachelor's degrees in physical science strongly correlates with the distance between Saturn and the moon during the period 2012–2021, see Tyler Vigen, Spurious Correlations, available at https://tylervigen.com/spurious-correlations

2.4 Causal Powers

When we talk about the cause of a certain effect, we are inclined to describe the situation as that some entity, a person, a physical object or a machine, has a certain *causal power*, which under certain conditions is manifested by bringing about the effect. It is, for example, common to say that the Earth has the power to attract bodies and that this explains why a stone falls to the ground. This causal power is the Earth's gravitation and this explanation appears to most people satisfying.

Another example: some persons are charming, which is a dispositional property of having the ability to charm other persons, thus sometimes causing certain behaviours in other persons. For example, one might explain a person's foolish behaviour by saying that he/she was charmed by another person, and the latter is well-known for their charming capacity. Thus, the terms 'dispositional property' and 'capacity' are often believed to express causal powers attributed to things.

One may first observe that a causal power is not the same as the cause of a particular event or action. The cause of a particular event is another event. But when we observe a regular connection between two types of events, for example that ponderable bodies fall to the ground when the support is removed, or that many people are charmed by a certain person, we are inclined to explain such a regularity by postulating a causal power in the entity thought to be the cause of this regularity. Thus, causal powers are invoked in causal *explanations* of recurring features and behaviour, although strictly speaking the causal power is not the direct cause of a singular effect.

Saying that a certain thing has a certain power is to ascribe to this thing a dispositional property, i.e., a property that only under certain conditions is manifested as an event that can be observed; but causal powers themselves are not observable, although thought to be permanent properties of things.

Postulating that an object has a certain causal power may be regarded as the proper explanation of our observations of some regular and recurrent behaviours, but it does not increase our ability to manipulate things or predict the future. This is so because any inference to future events is based on observed regularities, and these observed regularities are exactly the very reasons we have for inferring a causal power. Postulating a causal power does not increase the probability for a certain future event to occur. Hence, philosophers with an empiricist mind-set are sceptical about causal powers. The argument is Ockham's razor: 'do not without necessity postulate an entity.' (There are several formulations of this principle; the Latin version is 'Entia praeter necessitatem non esse multiplicanda.') We can make exactly the same inferences, with the same degree of certainty, to future events without causal powers.

Reflecting on Ockham's razor, one might ask 'necessary for what?' and it is pretty obvious that the tacit assumption is that the goal is to make predictions about events and states of affairs not observed when the utterance is made.

But causal powers are often held to have explanatory force, they give us under-standing of recurrent events, and one might be tempted to infer that understanding is a prerequisite for successful predictions about the future.

The validity of this line of thought depends crucially on the criteria for understanding a phenomenon. A prediction either succeeds or fails and that can be determined by observations. But what are the criteria for understanding? They seem to be crucially dependent on background knowledge had by those being given the explanation. We will discuss this topic further in Chap. 8.

Causal powers are unobservable, but that is not the relevant epistemological point. There are, for certain, many cases in the history of science where unobserved entities are postulated in order to explain observed phenomena. The crucial point is that such a postulated entity is accepted as real only when there is independent evidence for its existence. In the case of causal powers, there cannot be any such independent evidence; a causal power only manifests itself as a certain observed regularity, which is exactly the same as what is needed as the empirical basis for inferences about unobserved events.

Summarising, there is no empirical evidence for causal powers being responsible for the observed events in nature or for human behaviour. Nevertheless, many people hold that causal powers explain observable events. Whether that is so depends on our criteria for causal explanations.

2.5 Summary

Causal idiom is a basic feature of natural language, just as words for e.g., animals, colours, people, activities and events. The difference between on the one side words for these things, and on the other hand causal expressions, is that the latter concern relations between pairs of entities, while the former are talk about singular entities. In the second case, one observes *two* events and under certain conditions infers that they are related as cause and effect. Children learn causal expressions directly, in interactions with parents and other care-takers, not by being taught verbal definitions.

The notion of cause (and effect) can be expressed by quite a number of different words and expressions, for example, 'bring about', 'make happen', 'produce' and 'do'. None is more fundamental than the other.

When we ask for a cause, we have tacitly a certain event in mind, we ask for the cause of a particular event. The question has the form 'What is the cause of E?' Hence, the terms 'cause' and 'effect' are relational words; the effect (or the cause, if the effect is asked about) is often not mentioned as being obvious in the context at hand.

Questions about causes and effects are basically questions about relations between events and, secondarily, relations between types of events and their representations, quantitative and category variables. Casual relations between quantitative and category variables depend on causal relations between individual instances of these relations.

Discussion Questions

1. Is it possible to observe the causing *relation*, that a particular event *causes* another event? If your answer is yes, how, then, to distinguish between the mere succession in time of two events and them also being causally related?
2. Could a singular event be the cause of a category or vice versa?
3. Every day the sunrise is followed by sunset (provided you are not living in polar areas) without exception. Why not say that the sunset is caused by sunrise?
4. Many laws in physics are expressed as mathematical equations relating quantitative variables. Such laws, for example Newton's second law, ($force = mass{\cdot}acceleration$) is usually thought of as stating a causal relation. What is the justification for such a reading?

References

Grotzer T (2012) Learning causality in a complex world: understandings of consequence. Rowman and Littlefield Education, Lanham

Lindegren M, Mollmann C, Nielsen A, Stenseth NC (2009) Preventing the collapse of the baltic cod stock through an ecosystem-based management approach. Proc Natl Acad Sci 106(34):14722–14727

Pearl J (2000) Causality. Models, reasoning, inference. Cambridge University Press, Cambridge

Chapter 3
Causal Talk Is Fundamental

Abstract The terms 'cause' and 'effect' are very common in both ordinary and scientific discourse. Since they have a number of synonyms (or near synonyms), there is no point in trying to define 'cause' or 'effect' using any of these synonyms; 'cause' and 'effect' belong to the most fundamental level of language learnt in childhood. The way to give their meaning is to display a number of contexts in which causal expressions are used without any justification. The chapter ends with a presentation and discussion of Hume's criteria for the use of the term 'cause'. The main points are:

- Causal expressions belong to our most basic vocabulary, learnt at an early age.
- Many causal expressions from ordinary language are used in scientific texts without being explicitly defined.
- It is not possible to define 'cause' and 'effect' in any more primitive vocabulary.
- What we can do is to describe some necessary conditions for the correct use of 'cause' and 'effect'.

3.1 Introduction: The Pervasiveness of Causal Talk

The words 'cause', 'effect' and their synonyms are used mainly for two purposes: to explain things and to help us decide what to do in order to achieve a certain desired state of affairs. Therefore, it is important to have a deeper understanding of their meaning. One may view the meaning of a term as the rules for its correct use. This is not to say that there is anything like a well defined meaning of the term 'cause' (and its converse 'effect') that is the same in all contexts; as with very many expressions, the meaning depends to some extent on context.

As observed in the previous chapter, relating events or states of affairs as cause and effect is a basic feature of our thinking, acting and use of language. This fact strongly suggests that it is not possible to define the term 'cause' (or its converse 'effect') using more primitive vocabulary; we learn to use the words 'cause' and 'effect' when we learn to talk, without being given any verbal explanations. We do not first learn a basic vocabulary that later is used in an explicit definition of

'cause'. What we instead can do is to state a number of conditions for the correct use of 'cause', for example, that in a singular case a cause never happens later than its effect.

Causal claims are often made without any scientific backing. As an illustration, look at the following list of quotes, randomly selected, from New York Times 2020-08-17, where 'cause' and 'effect' are italicised by us, not by NYT:

'But when the pandemic *caused* demand for bikes to jump, Giant needed to reverse course.'

'Gov. Gavin Newsom of California called for an investigation into what he described as a major utility failure that was even more alarming set against the backdrop of the pandemic, when people, many largely confined inside, may be more dependent than ever on electricity: rolling blackouts over the weekend, *caused* by a record-shattering heat wave.'

'Problems with coronavirus data reporting *cause* confusion in Iowa and beyond.'

'Uncertain still was what *effect*, if any, the event would have on the spread of the virus.'

'The Interstate Highway System was justified in part with dubious claims about national security, but it had the *effect* of reinforcing national unity.'

'Dr. Kimberly Manning, an internal medicine doctor at Grady Memorial Hospital in Atlanta, recalled countless micro-aggressions in clinical settings. "People might not realise you're offended, but it's like death by a thousand paper cuts," Dr. Manning said. "It can *cause* you to shrink."'

These causal claims (understandably, several relating to the Covid-19 pandemic!) are not explicitly backed by any profound scientific analysis; the NYT journalists and the persons' referred to take the readers to understand and accept the causal claims as correct against a background of common knowledge.

The scientific literature is no less filled with the words 'cause' and 'effect'. In a paper discussing Bertrand Russell's thesis that there are no causal laws (Russell, 1913), the authors started to look for the prevalence of the words 'cause' and 'effect' in a leading scientific journal:

> A search for articles in which the word 'cause' appeared in the on-line archives of *Science* between October 1995 and June 2003 returned a list of results containing 8288 documents, averaging around 90 documents per month, in which the word 'cause' occurred. 'Effect' was more popular - 10456 documents for the same period, around 112 per month. (Ross and Spurrett, 2007, 60)

But we use not only the words 'cause' and 'effect' for talking about causation; expressions such as 'bring about', 'lead to', 'make happen', 'produce', 'result in', do', etc., also express causal relations. And we have also the negative counterparts to these, such as 'stop', 'prevent' and 'hinder', expressions that suggest causal actions performed in order to bring an end to some undesirable state.

The same is true of 'effect', which in many contexts can be replaced by 'result', 'outcome', 'final state', 'impact', etc. So we submit that if the authors of the paper had included these other causal expressions indicated above, the figures would have been still higher. It is safe to say that causal thinking is deeply ingrained in our thinking, it belongs to our nature.

That causal thinking is a fundamental aspect of our being as humans was a point already made by Kant in his *Critique of Pure Reason* (Kant et al., 2003, (A80/B106)). He argued that our mind has certain structural features that determine the forms of our judgements. Thus, Kant identified 12 fundamental concepts, called *categories*[1] among which *cause* is one. These categories make up the conceptual basis for all judgements, according to Kant.

His arguments are convoluted and much debated, but one may arrive at a similar conclusion, at least in the case of causal judgements, without taking any stance about his transcendental arguments. One may, as was done above, observe ordinary language use, instead of speculating about our minds. Doing that, we realise how common and basic the use of causal expressions are in ordinary discourse.

3.2 Attempts to Define 'Cause'

In science we are required to define central terms. Giving a verbal definition of a term is to give its meaning in terms of more common and well known expressions. This is hardly possible when it comes to such a basic term as 'cause'.

At the online Merriam-Webster dictionary we are given the following list of synonyms for 'cause' used as verb: 'beget', 'breed', 'bring', 'bring about', 'bring on', 'catalyse', 'create', 'do', 'draw on', 'effect', 'effectuate', 'engender', 'generate', 'induce', 'invoke', 'make', 'occasion', 'produce', 'prompt', 'result (in)', 'spawn', 'translate (into)', 'work' and 'yield'.

If we attempt to define 'cause' in terms of e.g., 'bring about', we might further ask about the meaning of 'bring about'. Then a natural response would be to say that it means 'cause'. In common language, none of these terms is more basic and informative than the others.

In Nancy Cartwright's *How the Laws of Physics Lie* we read:

> Causes make their effects happen. We begin with a phenomenon which, relative to our other general beliefs, we think would not occur unless something peculiar brought it about. (Cartwright, 1983, p. 76)

Here Cartwright gives the meaning of 'cause' in two ways: first as synonymous to 'make happen', then using a counterfactual, a sentence of the form 'If A had not occurred, B would not have happened.' The first would suffice if the defined term is less well understood than the defining term, but that is hardly the case with 'cause' and 'make happen', as earlier pointed out. Why not reverse and explain the meaning of 'make happen' in terms of causation? The term 'cause' is one of the most common and basic words in natural language. So maybe that is why she immediately moves to a counterfactual explanation.

---

[1] Kant's use of the term 'category' is quite different from our modern term, see the glossary item 'category variable'.

Cartwright is not alone in explaining the meaning of the term 'cause' by a counterfactual expression, it is quite common in philosophy. But how do we know that a counterfactual statement is true? Obviously it can never be justified by an observation. This will be further discussed in Chap. 4.

The notion that we somehow could define 'cause' and its converse 'effect' using a more basic and clearer vocabulary seems indeed dubious. The meanings of 'cause' and 'effect' and their synonyms are determined by the rules we automatically and without any conscious justification apply when we use these terms in concrete communication situations. We talk about causes and effects mainly for two purposes; to explain things and to decide what to do in order to achieve our goals; successful doing is causing a desired event to happen. Causal thinking and causal talk is a basic trait of humans, which, by the way, also was Kant's conclusion.

We have in ordinary language no explicit criteria of application for the terms 'cause' and 'effect'. But when they are used in a scientific context, one must state criteria in order to know what can be inferred from causal statements. Thus Cartwright writes:

> Like Machamer et al. (2000) I too have long followed Anscombe's view that the ordinary concept of 'cause' is highly general. It is what, following Otto Neurath, I call a 'Ballung' concept. A Ballung concept is a concept with rough, shifting, porous boundaries, a congestion of different ideas and implications that can in various combinations be brought into focus for different purposes and in different contexts. Many of our ordinary concepts of everyday life are just like this. Ballung concepts also can, and often do, play a central role in science and especially in social science. But they cannot do so in their original form. To function properly in a scientific context they need to be made more precise. This will be done in different ways in different scientific sub-disciplines, serving different ends and to fit with the different concepts, methods, assumptions, and standards operating in these disciplines. The more precise scientific concepts that result will in general then be very different from each other and different yet again from the original Ballung concept. (Cartwright, 2017, 136)

We basically agree with Cartwright that in scientific contexts we need to clearly state the conditions for the legitimate use of 'cause', 'effect' and their near synonyms.

3.3   Are Causal Connections Observable?

Given the omnipresence of the term 'cause' (and 'effect') in everyday life and in science, what, then, are the criteria for it and its synonyms? Can one directly observe a causal connection between two events?

Some philosophers (for example G.E.M. Anscombe (1971)) argue that we can, in certain cases, observe causation. We are not convinced. One may reasonably be doubtful on the ground that 'cause' and 'effect' are relational terms; the basic syntax is 'x is the cause of y' and 'y is the effect of x', where the placeholders 'x' and 'y' represent events, states of affairs, or aggregates of such things described by categories or quantitative variables. We observe events and states of affairs, but do

we observe their causal connection? Do we really *see* or *hear* the cause propagating its impact to the effect? Can one really say that we directly observe any kind of connection of any two things? We think not. Perception is primarily perception of objects and, derivatively, of their changes.

Empiricists are prone to say the same, they restrict the use of the term 'direct observation' to things we discern with our sense organs.

But how to draw the line between a direct observation and an inference from such an observation? Consider the following situation: you enter a new hotel room and want to turn on the light. You push the first button you see and the bathroom, not the room, is lit. Then you push another button and the room is lit. Can one say that you observed that pressing the first button caused the bathroom to be lit, pressing the second one caused the room to be lit? Would a person who never before has been in a modern building with electric lights, for example, a person belonging to a hunter-gatherer culture in the Amazon rainforest, say that they saw the causing? We think not.

When people are prone to say that they observed these two events being causally connected, they rely on previous experience and tacit inferences made from such experiences, and some knowledge about electricity. In general, inferences to causal relations between events are based on experiences from experimentation; if we manipulate one object in certain ways and observe changes in some other object so that one can control the states of the second object by doing things with the first, we apply the cause-effect relation. And this is an inference, not a direct perception.

If we, on the contrary, describe this situation as a direct observation of a cause-effect relation, we have in fact made the well known fallacy *post hoc, ergo propter hoc*.[2] We distinguish between mere succession and causation and that cannot be done by mere observations; ultimately, we must perform experiments by manipulating one variable and observing the other. A statement about a causal relation is the result of an explicit or implicit inference from such experiments.

3.4 Hume's Criteria for the Use of 'Cause'

David Hume discussed the observational basis for talk about causation (Hume, 1986/1739). His proposal was that there are basically two directly observable features of a pair of events that trigger (i.e., *cause*!) us to say that they are related as cause and effect:

1. The cause precedes its effect.
2. Cause and effect are contiguous.

But this is obviously not sufficient. There are many cases of pairs of events/states of affairs being in physical contact and one of them preceding the other, without us

---

[2] 'After this, hence because of this.'

saying that the first one is the cause of the other. Hume therefore added the *regularity condition*, popularly stated as 'same cause, same effect'. Expressed more carefully: A is the cause of B if and only if A belongs to a type of events/states of affairs that regularly is followed by another type to which B belongs, and if A and B satisfy the other two conditions.

Thus Hume's analysis of the use of 'cause' and 'effect' may be summarised as that each of the three conditions: (1) the cause precedes its effect, (2) cause and effect are contiguous and (3) the same type of cause is regularly followed by the same type of effect, are necessary and that they are jointly sufficient for the correct use of sentences of the form 'x causes y'.

Each of the conditions has been doubted and Hume in fact discussed caveats to all three (Hume, 1986/1739). Regarding timing, he accepted that cause and effect sometimes could be simultaneous. Regarding contiguity he realised that there could be intermediate events/states of affairs so that cause and effect may be indirectly connected via a chain of intermediate events/states of affairs. The causal relation is transitive. Finally, about regularity, he accepted the possibility of several causes for a particular effect, in which case we must say that a particular cause is not always followed by its effect, for other causes may also be needed. So a particular cause only increases the probability for the effect to occur.

Many philosophers have been critical towards Hume's regularity theory, the main argument being that it does not really explain what a cause is. Many people ask for explanations of regularities, usually in terms of causal powers. For an empiricist this is reversing the order of explanation; if there are any such things as causal powers, these must be explained in terms of observations, i.e., observed regularities, see Sect. 2.5.

But there is another problem with the regularity view: there are many regularities that we do not count as instances of causation. It is a well established piece of knowledge that correlation is not causation. How to distinguish cases of correlation that indicates causation from those that do not?

The first step is to use Hume's condition 2; individual instances of causes and effects must be in contact, provided we can give a clear meaning to the notion of contact. The problem is that events, states of affairs, properties, or other kinds of entities being related as cause and effect hardly can be said to be in contact, since they are not bodies.

It is easy to grasp the underlying idea that there must be a process, some kind of physical, chemical or biological link between cause and effect, when these are two individual events in space and time. But how to apply that to, e.g., a causal relation between two attributes?

One hint may be found by reflecting on how we distinguish between correlations and causal relations. An observed correlation between two variables, i.e., events/states of affairs of types A and B respectively, is a causal relation only if there is a mechanism, a chain of events/states of affairs connected by physical signals (remember: 'physical' here includes chemical and biological events) transferring information from any particular A-event to a particular B-event. And how do we know that there was a transfer from an A-event to a B-event and not the opposite?

There are two factors determining this: timing and deliberate manipulation of the A-event.

The contact requirement is a *necessary condition*. This is illustrated by the fact that we do not believe in extra-sensory perception (ESP). Some people claim that they can acquire information about things from which no physical, chemical or biological signals could have reached their mind, but no evidence has been produced. Several well-conducted experiments performed with persons claiming to have extra-sensory capacities have been made and they have all failed; there is simply no empirical evidence for ESP. Perception requires physical signals triggering our sense organs, so if there is no physical signal from an individual A-event to an individual B-event, the A-event cannot be a cause of the B-event.[3]

These reflections indicate, again, that experiments where one variable is manipulated and another is observed are crucial for establishing a causal relation. If we observe a variation in an observed variable following variations in a variable being manipulated, we infer that there is a causal link connecting the two variables. And we take it for granted that there is a physical, chemical or biological mechanism making up the connection between pairs of singular events.

So causal relations between variables, quantities and other abstract things are grounded on causal relations between those individual events making up these abstract entities; physical signals transferring information between individual events or states of affairs make perfect sense and has been discussed by several philosophers, e.g. Reichenbach and Reichenbach (1999), Salmon (1984), Salmon (1997), Salmon (2001), and Collier (1999).

This is not to say that physical links between cause and effect always is a salient aspect of a causal explanation. Talking about causes of historical events, wars for example, the physical connections between power centres, (letters or telegrams sent between presidents and prime ministers before a war) are rarely of any relevance for our questions about causes in history. But there must be such links.

The fundamental method to obtain information about physical links is to perform experiments, to be further discussed in Chap. 7. But it all depends on our ability to keep factors other than the hypothesised cause under control. How do we ascertain that, when we do not know which other factors there might be? That is often our problem when trying to decide which causal connections there are between parts of a complex social, ecological or social-ecological system.

Another way to decide whether a correlation between A and B is due to a causal link or not is to use earlier established scientific theory to describe causal mechanisms, to be discussed in Chap. 8. But this earlier established theory must in turn be based on experimental evidence for causal links.

---

[3] In particle physics there is a phenomenon called 'entanglement' which seems to show that information without signals is transferred between distant objects. This is however a misrepresentation of the facts, see e.g. Johansson (2007).

3.5   Summary

Our fundamental causal terms, ('cause', 'effect', 'make happen', 'bring about', etc.) belong to our basic vocabulary, learnt as part of learning ones mother tongue. This means that one cannot define 'cause', 'effect', etc., in some more basic vocabulary. The meaning of such words are learnt by learning their application to a number of concrete situations as experienced by the child.

Cause and effect are relational terms, they are predicates of the form '.... is the cause of....' and '.... is the effect of.....'. We cannot directly observe causal relations; we observe physical objects and events, and under certain conditions two such observed events satisfy the predicate '...is the cause of......'

The conditions for saying that an event is a cause of another event was first formulated by Hume. They are (1) cause precedes its effect (or is practically simultaneous), (2) cause and effect are in contact (directly of indirectly) and (3) same types of causes are regularly followed by same type of effects. All three conditions have been extensively discussed by philosophers. We hold that Hume was almost correct. The only improvement needed is that the regularity condition should be restricted, namely, that the correlation ('regularity') is observed in experiments where the cause is manipulated and the effect varies accordingly. Hume's three conditions might have been a correct description of common use of the term 'cause' in his days, but we have later learnt to be more restrictive.

Discussion Questions

1. Do you accept our conclusion that in ordinary language there is a rather big set of expressions each being roughly synonymous with 'cause' and such that none is more basic or more easily understood than the other in this set? If not, what is your idea about the most fundamental term? Why this choice?
2. One suggested definition of 'cause' goes: 'A cause is that what makes its effect happen'. Do you think this is a satisfactory definition or not?
3. Do you agree with Hume (and us!) that it is impossible that the effect precedes its cause? Why?
4. In many cases we do not know the physical, chemical or biological mechanism connecting a cause and its effect, nevertheless we are convinced that the causal link exists. Give an example of such a situation and explain the reason why a causal connection nevertheless is believed to exist.
5. How do social mechanisms differ from physical, chemical or biological mechanisms?

References

Anscombe GEM (1971) Causality and determination: an inaugural lecture. Cambridge University Press, London

Cartwright N (1983) How the laws of physics lie. Oxford University Press, Oxford

Cartwright N (2017) Can structural equations explain how mechanisms explain? In: Beebee H, Hitchcock C, Price H (eds) Making a difference: essays on the philosophy of causation. Oxford University Press, Oxford, Chap. 8

Collier JD (1999) Causation is the transfer of information. In: Sankey H (ed) Causation and laws of nature. Kluwer Academic Publishers, Dordrecht, pp 215–246

Hume D (1986/1739) Treatise on human understanding. Penguin, London

Johansson LG (2007) Interpreting quantum mechanics. A realist view in Schrödinger's Vein. Ashgate, Aldershot

Kant I, Timmermann J, Klemme, HF (2003) Die Drei Kritiken. Kritik der Reinen Vernunft. Meiner, Hamburg

Machamer P, Darden L, Craver C (2000) Thinking about mechanisms. Philos Sci 57:1–25

Reichenbach H, Reichenbach M (1999) The direction of time. Dover, Mineola

Ross D, Spurrett D (2007) Notions of cause: Russell's thesis revisited. Br J Philos Sci 58(1):45–76

Russell B (1913) On the notion of cause. Proc Aristotelian Soc 13:1–26

Salmon WC (1984) Scientific explanation and the causal structure of the world. Princeton University Press, Princeton

Salmon W (1997) Causality and explanation: a reply to two critiques. Philos Sci 64(3):461–477

Salmon W (2001) A realist account of causation. In: Marsonet M (ed) The problem of realism. Ashgate, Burlington, pp 106–133

Chapter 4
Causation, Counterfactual Dependence and Potential Outcomes

Abstract When talking about causes we often think of an imagined contrast to the real sequence of events: we use a counterfactual, asking what *would* have happened if the cause had not occurred. But one might be doubtful about the explanatory force of this analysis. The basic problem is that the truth or falsity of a counterfactual statement cannot be determined by empirical means. In some cases, notably in physics, we can apply a strict law when justifying claims about alternative scenarios. In most cases, however, we have at best regularities and in such cases it is difficult to have any confidence in the corresponding counterfactual. An account of causation in terms of the more restricted concept of potential outcomes is much more useful. It is closer to empirical practice and is more reliable. The main points are:

- It is common to explain the meaning of A caused B' as 'If A had not occurred, then B would not have occurred.'
- This is no step forward, since we do not in general know whether the counterfactual is true.
- Rubin's more restricted notion of potential outcome is to be preferred, since it can be applied to empirical research.

4.1 Introduction

The concept of cause (as used in sentences of the form 'A caused B', or its converse 'B is an effect of A') is quite often explained as 'If A had not been the case (or occurred), then B would not have been the case (or happened)'. It is a counterfactual (or contrary-to-fact) statement since A and B in fact both occurred.

How, then, do we know what *would* have happened if A had in fact not happened? It is obvious that our justification for saying what would have happened, if the stream of events had differed from what actually occurred, must be based on some inference from actually occurring and observed regularities to unobserved events. How do we ascertain that?

This is an instance of the problem of induction, the problem to state under what conditions an inductive inference can be relied upon.[1] If we have a trustworthy and exception-less law at our hand, the problem is solved. Instead of using observed states as initial conditions in calculating future states of affairs, we can put in non-actual values and use the strict law to calculate what would have happened in such a non-actual case. But in most situations we have no strict law at our disposal. We will further discuss the connections between laws and causation in Chap. 7.

This will not work if we use non-strict laws, i.e., laws with so called ceteris paribus clauses ('All else being the same'). This is so because when we imagine a non-actual initial condition we do not know whether other relevant circumstances also would differ from the actual observed situation. When we suspect that our law is not strict we add the clause 'ceteris paribus' just because we do not know which factors we need to take into account or not.

Can one say something about counterfactuals and causation without using strict laws? Empiricists are sceptical. Stating truth conditions for counterfactuals has proven to be a deep problem and there is no agreement about its solution. It appears, if anything, more difficult than stating truth conditions for statements about causal relations.

4.2   Goodman on Counterfactuals

The seminal paper in the discussion about counterfactuals was Nelson Goodman's *The Problem of Counterfactual Conditionals* (Goodman, 1946). In that paper Goodman observed that there is a profound semantic difference between counterfactual and indicative conditionals.

Counterfactual and indicative conditionals differ in verb form; counterfactuals are expressed using the subjunctive mood, whereas indicative conditionals are expressed in indicative mood, and this difference indicates a semantic difference. The following example may illustrate:

 Indicative conditional: If it is raining right now, then Sally is inside.
 Counterfactual conditional: If it were raining right now, then Sally would be inside.

---

[1] This is the modern conception of the problem of induction, emanating from Goodman. The traditional problem was to give a general justification of inductive reasoning, which proved impossible.

These two sentences differ in meaning, hence there is a difference in their truth conditions. In order to analyse this difference, we start with the truth table for the indicative conditional (also called the material conditional):

A	B	If A, then B
True	True	True
True	False	False
False	True	True
False	False	True

We can formulate the content of this table as that an indicative conditional is true whenever the antecedent is false or the consequent is true.[2]

If we apply this truth table to the counterfactual conditional it will come out true, since the verb form 'were' means that what follows is in fact not the case. Hence 'were' tells us that the antecedent is false, it is not raining. So far, so good. But according to the truth table it doesn't matter whether Sally is inside or not, both #1 and #2 come out true:

#1. If it were raining right now, then Sally would be inside.
#2. If it were raining right now, then Sally would not be inside.

That cannot be correct, they contradict each other. Hence, the truth conditional analysis where antecedent and consequent are evaluated separately must be wrong in the case of counterfactual statements. No truth table can account for the semantics of counterfactual sentences.

As Goodman observed, it is some sort of connection between the events described in the antecedent and the consequent that determines the truth value of a counterfactual statement. This connection is not, and cannot be, reflected in any *merely logical* connection between antecedent and consequent, since logic concerns the *forms of sentences* and *formal relations* between sentences.

Goodman next observed that the difference between true and false counterfactuals is that true counterfactuals are connected to laws, while false are connected to true, *accidental generalisations*. Here is one example of a contrast between a law and an accidental generalisation (not discussed by Goodman, but by several later philosophers):

3. All spheres of gold are less than 1 km in diameter.
4. All spheres of U-235 are less than 1 km in diameter.

#3 and #4 have the same logical form and presumably are they both true. (If one would find somewhere in the universe a really big lump of gold, we could have a

---

[2] Indicative conditionals are used for making conditional claims: the consequence is claimed to be true under the condition that the antecedent is true. If the antecedent is false, no claim about the consequent is made, it may be true or false.

longer diameter.) But there is a difference; #4 is a consequence of fundamental laws of nuclear physics, and therefore itself a law, which means that it is not merely true but necessarily so; it is impossible to assemble an amount of U-235 bigger than the critical mass (52 kg, a sphere with diameter = 17 cm) of this radioactive isotope.[3] By contrast, #3 is accidentally true; it just happens to be no big lumps of gold in the universe.

Based on #3 and #4 we can now construct two counterfactuals, one true and one false:

#5. If x were a sphere of gold, it would be less than 1 km i diameter.
#6. If x were a sphere of U-235, it would be less than 1 km in diameter.

As far as we know, #5 is false while #6 is certainly true. So it seems reasonable to say that true counterfactuals are based on laws whereas false counterfactuals are based on accidental generalisations.

This would be a real step forward, if we had a clear explanation of the difference between laws and accidental generalisations. But we have not; many people have strong intuitions about the difference, but so far no generally accepted analysis is in sight.[4]

Goodman concluded his paper by admitting that he had no solution to the problem with counterfactual conditionals, since he had no suggestion of how to distinguish between laws and accidental generalisations.

The discussion about causation and counterfactuals has been intense and one may discern two main strategies: either to analyse causation in terms of counterfactuals, or the other way round. This choice is guided by ones metaphysical views: David Lewis (1973) and many others think that a semantics of counterfactuals in terms of possible worlds is satisfactory and taking that as a firm ground one can then define causation in terms of counterfactuals. Those sceptical about the existence of possible worlds, or even the intelligibility of this notion, (How do you identify a possible but not actual world?) hold that the explanation should go from causation to counterfactuals.

We belong to this latter camp. Counterfactual statements belong both to our vernacular and to scientific discourse; they are widely used and there is no reason to assume that users of this idiom tacitly or explicitly delve into deep metaphysics concerning possible worlds and our access to them. Hence, we think counterfactuals should be explained in terms of more basic concepts such as causes or perhaps laws. As we showed in Chap. 3, causal talk belongs to our very basic vocabulary, learned already when first learning to talk our mother tongue. This means that explanations of the meanings of less basic expressions should be done in terms of the basic vocabulary.

---

[3] This impossibility is a consequence of fundamental properties of U-235 nuclei, which can be derived from quantum theory.

[4] But see Johansson (2019) for an empiricist account of strict laws not built upon counterfactuals or necessity.

In 1940s, when Goodman's paper was published, there were little discussion about the concept of a law of nature. The received view was that when a hypothetical general statement was supported by a sufficient number of observations and no counter instances were observed, one had reason to believe that the general conclusion was correct, and it was then elevated to being a law. Those who elaborated the details of this line of thought argued that probability arguments could be used. But this idea met, justifiably, devastating criticism. Having high probability is not the same as certainty, and strict laws are certain. Furthermore, as Goodman pointed out, there is a profound difference between laws, which are necessary, and other true general statements of the same logical form, which are not necessary, and this difference could not be analysed using only empirical arguments.

What does this mean for the counterfactual analysis of causation? Our view is that in so far as we are unclear about the meaning of 'cause', giving this concept a definition in terms of counterfactuals is no step forward; counterfactuals are strongly related to laws, and both the notions of *counterfactual* and *law* are less clear than that of *cause*. So what to do? James Woodward has suggested a way out.

4.3   Woodward's Account of Causation

James Woodward has discussed the counterfactual analysis of causation in several papers (Woodward, 1997, 2002, 2003, 2008, 2016). One might guess that Woodward was inspired by Goodman's observation of the strong connection between true counterfactuals and laws, although he made no references to Goodman's paper. Taking into account that the concept of a law of nature is as much in dispute as are counterfactuals, Woodward's step forward was to base true counterfactuals on what he called 'invariances'.

An invariance is an observed regularity, although not one elevated to the status of being a law. Thus Woodward was able to avoid the metaphysical jungle of necessities. Neither is an invariance merely an accidental generalisation. Woodward's idea is that an observed regularity which has been used in several successful predictions may be labelled an *invariance*, which tells us that it is a weaker concept than that of a natural law. But what, more precisely, is the difference?

Woodward intended 'invariances' to refer to regularly occurring phenomena restricted to some region in space and time. One could for example say that it is an invariance (or 'restricted regularity') that almost all people who has spent 10 years or more in Sweden understand, to some degree, Swedish, while hesitating to call this regularity a 'law'. But it depends on what we mean by 'law'.

In any case, if we accept this regularity we are prone to accept as true the counterfactual 'If NN had been living in Sweden for 15 years, she would understand Swedish', said about a certain person that only understands her mother tongue, say Swahili, and has never been in Sweden.

This type of local and restricted invariances differ from laws in that they are not exception-less. Observing such an exception we are prone to ask for an explanation, i.e., a causal explanation. We are back to causes.

One further difference between laws and invariances is that laws properly so called are integrated into a theory consisting of a number of laws logically related to each other.

The question is: can one refer to an invariance as evidence for a counterfactual? It seems that the answer is no, unless we know the causal mechanism producing the invariance. For how could an invariance observed in a number of cases be known to be valid also in a non-observed case? Invariances may have exceptions, they are not strict laws, and how do we know that an unobserved case is not an exception?

It seems that an 'explanation' of the concept of cause in terms of counterfactual dependence is no step forward. It is much more reasonable to say that we can explain 'counterfactual dependence' in terms of causes. The word 'cause' and it's synonyms ('bring about', 'lead to', 'produce', etc.) belong to common language and is much easier to understand than any technical term.

4.4 Potential Outcomes Instead of Counterfactuals?

Instead of analysing causation in terms of counterfactuals, Rubin, following Neyman (1923) and Fisher (1925), uses the concept of *potential outcomes*. Here is how he motivates it:

> Some authors (e.g. Greenland et al., 1999; Dawid, 2000) call the potential outcomes "counterfactuals", borrowing the term from philosophy (e.g. Lewis, 1973). I much prefer Neyman's implied term 'potential outcomes' because these values are not counterfactual until after treatments are assigned, and calling all potential outcomes 'counterfactuals' certainly confuses quantities that can never be observed (e.g. your height at the age of 3 if you were born in the Arctic) and so are truly a priori counterfactual, with unobserved potential outcomes that are not a priori counterfactual (see Frangakis and Rubin (2002), Rubin (2004); and the discussion and reply for more on this point.) (Rubin, 2005, 325)

Here is a simple illustration of how to use the concept of potential outcomes. Suppose we have randomly divided a test sample, taken from some population, into two groups, one consisting of those being treated in some way, the rest is the control group. For each unit in the treatment group one can only observe its actual state after the treatment and similarly for the control group; in this group one can only observe the state of a unit after not being treated during the experiment. Both the actual state and the non-actual possible state of a unit are *observable*, although only one state is actually observed. Hence the term 'potential outcome'. We can now compare the observed outcomes in the two groups. We can calculate the conditional probability for the outcome B, conditioned on the intervention A, $p(B|A)$ and compare with the marginal probability $p(B)$. If $\text{prob}(B|A) > \text{prob}(B)$, we have strong reason to believe that A is a cause of B. (N.B. the indefinite 'a cause'; there may be more causes!) The

intervention A may be a intentional manipulation or an intervention not planned by the experimenter, i.e., a so called 'natural experiment'.

Replacing the concept of potential outcome for counterfactual in discussions about causation is a significant step towards a more empirical approach. Moreover, it connects to the manipulability account of causation, see Sects. 3.4 and 6.1. It is useful when making inferences about causation from observed results of experiments, and also in making inferences from so called 'natural experiments', see Sect. 6.2.

From a philosophical point of view this is a significant improvement as compared with the counterfactual analysis. The semantics of counterfactuals in terms of possible worlds faces two obstacles: (1) how do we identify a possible world and (2) *which* possible worlds should we take into account when describing the semantics of causation? One needs to impose restrictions on what to count as a possible world, which is usually made in terms of similarity to our actual world. This can be made formally stringent, but it is not helpful for the empirical researcher, since it leaves the notion of similarity with the actual world undefined. The crucial question is 'Similar in what respect?'

The terminology of potential outcomes is, in comparison, applicable to actual experiments and observations. The set of potential outcomes are defined in the experimental design. We perform an experiment and explicitly state the set of possible outcomes, of which one is actualised. For further discussion see e.g., (Menzies and Beebee, 2020) and (Rubin, 2005).

4.5   Summary

In ordinary parlance we take it for granted that the sentence 'A is the cause of B' is more or less synonymous with 'If A had not occurred, B would not have occurred.' This assumption is then used for explaining causation in terms of counterfactuals. But on second thoughts one may reasonably conclude that this is not of much value as an explanation; the meaning of counterfactuals is much more foggy than the meaning of 'cause'. How do we know what would have happened if the course of events had been different from what actually happened?

It seems that only if we have a strict scientific law at our disposal can we know with some certainty what would have happened, if the conditions had been different from what they actually were. But in very many cases we know no strict laws, so an analysis of causation in terms of counterfactuals is no step forward.

Woodward has suggested to use term 'invariance' instead of laws for explaining causation; Invariances are inferred from observed regularities in some local setting and believed to be true also in unobserved cases in the same type of settings, while not being elevated to the status of law and not integrated into a theoretical structure of laws related to each other.

A somewhat similar approach is taken by e.g. Rubin, who has suggested replacing the concept of *potential outcome* for *counterfactual*. Given a dynamical

equation we can calculate what would be the outcome for any chosen initial condition, actual or not. This equation guides the time evolution from the initial situation and we can map the set of selected initial conditions onto the set of potential outcomes. The dynamics may be strict (mapping one initial state onto one final state), or probabilistic, mapping a set of possible outcomes from each initial state). The list of alternative outcomes are clearly stated and all are observable; but only one will be observed, that which is actualised.

Discussion Questions

1. The truth table for the indicative conditional (displayed in Sect. 4.2.) are by many students not immediately accepted. Many often wonder about the combinations where the antecedent is false. Why is it correct to say that a conditional statement is true whenever the antecedent is false?
2. It is not possible to construct a truth table for the counterfactual conditional. Why?
3. Are there any laws that are not necessary? If there is any such example, why, then, is it said to be a law?
4. What is the difference between an accidental generalisation (a 'regularity') and an 'invariance' in the sense of Woodward?
5. Why is an analysis of causes in terms of potential outcomes to be preferred over an analysis in terms of counterfactuals?

References

Dawid AP (2000) Causal inference without counterfactuals. J Am Stat Assoc 95:407–448

Fisher RA (1925) Statistical methods for research workers. Oliver and Boyd, Edinburgh

Frangakis CF, Rubin DB (2002) Principal stratification in causal inference. Biometrics 58:21–29

Goodman N (1946) The problem of counterfactual conditionals. J Philos 44:113–128

Greenland S, Pearl J, Robins JM (1999) Causal diagrams for epidemiologic research. Epidemiology 1(10):37–48

Johansson LG (2019) An empiricist view on laws, quantities and physical necessity. Theoria 85(2):69–101

Lewis D (1973) Counterfactuals. Blackwell, Oxford

Menzies P, Beebee H (2020) Counterfactual theories of causation. In: Zalta EN (ed) The Stanford Encyclopedia of Philosophy, winter 2020 edn. Metaphysics Research Lab, Stanford University

Neyman J (1923) On the application of probability theory to agricultural experiments. Essay on principles. Section 9. Roszniki Nauk Rolniczych 10:1–51

Rubin DR (2004) Direct and indirect causal effects via potential outcomes (with discussion). Scand J Stat 31:161–179, 195–198

Rubin DB (2005) Causal inference using potential outcomes: design, modelling, decisions. J Am Stat Assoc 100(469):322–331

Woodward J (1997) Explanation, invariance, and intervention. Philo Sci 64(Supplement. Proceedings of the 1996 Biennal Meeting of the Philosophy of Science Association. Part II. Symposia Papers):S26–S41

Woodward James (2002) What is a mechanism? A counterfactual account. Philos Sci 69:S366–S377

Woodward J (2003) Making things happen: a theory of causal explanation. Oxford University
    Press, Oxford
Woodward J (2008) Invariance, modularity, and all that: Cartwright on causation. In: Hartmann S,
    Hoefer C, Bovens L (eds) Nancy Cartwright's philosophy of science. Routledge, London
Woodward J (2016) Causation and manipulability. In: Zalta EN (ed) The Stanford Encyclopedia
    of Philosophy, winter 2016 edn. Metaphysics Research Lab, Stanford University

Part II
Causation in Science

Chapter 5
Causal Relations and Causal Relata in Science

Abstract There is no such class of things as *causes*. The terms 'cause' and 'effect' are relational terms, in other words, our concern is the *causal relation* 'x causes y'. This relation is applied to pairs of singular events and states of affairs, to pairs of types of events/and states of affairs and to pairs of variables. In science we are mostly interested in generalities, so the focus is usually how to infer causal relations between types of events/states of affairs and between variables.

The main points in this chapter are:

- Both singular and general events/states of affairs can be related as cause and effect.
- Expressions for laws, equations, correlations and conditional probabilities do not by themselves state any causal relations. A causal interpretation of these requires information not contained in that law, equation, correlation or probability statement.
- Without background information it is impossible to infer a causal relation about a singular case, no matter whether the data are quantitative or qualitative.
- Most events have many causes, and such situations can be analysed in terms of INUS-conditions.

5.1 Introduction

Implicit in all talk about causes is the prior identification of an event, or a type of events, called 'the effect'. When we for example ask about the cause of the global warming, it is the global warming that is the effect. This remark may seem utterly trivial, but the point is that causes do not make up a distinct category of things, events or states of affairs. Discussions about causes are discussions about *causal relations*.

Statements about relations between two entities have the canonical form 'xRy', where 'R' is short for a two-place predicate. The things related are called 'relata'. In the case of causal relations it is 'x is a cause of y', or 'y is an effect of x'.

Relations relate things, objects, events, states of affairs, properties, variables and perhaps other things as well. From both an ontological and an epistemological point of view it is important to distinguish between singular and general causation.

5.2   Singular Versus General Causation

5.2.1   Causal Relations Between Singular Events/States of Affairs

The basic use of causal expressions relates, singular events; one event is said to cause another event. One might be uncertain which particular event was the cause of some identified event; still the basic idea is that a certain event, called 'the effect', is caused by another event and the effect in turn causes one or several other events. Thus, causal relations are transitive.[1]

An illustration from the present pandemic: when a particular individual A is infected by Covid-virus, we know that the cause is that he/she had been in too short a distance from another infected person B; this event, A and B coming into close vicinity to each other, is the immediate cause of A being infected. Hence the first and most obvious measure for diminishing the spread of Covid is that people keep distance.

It is also part of ordinary and scientific language to express causal relations between states of affairs. The state of affairs that the temperature in Stockholm was below zero for some weeks in January 2021 caused the lakes in this area to be frozen. These frozen lakes were for some time a state of affairs and it was the effect of the low temperature.

Perhaps there is no sharp distinction between events and states of affairs, and for the purpose of analysing causation it doesn't matter. For the present purpose it suffices to observe that events and states of affairs are individual things, i.e. entities referred to by singular terms,[2] which may be related as cause and effect. This is the basic form of application of the two-place predicate '... causes...'.

But more things than individual events and states of affairs have since long been said to be related as cause and effect. The first extension is to *types* of events and states of affairs.

---

[1] Some authors, e.g., Menzies (2012), hold that causation is not always transitive. The purported examples of non-transitive causation have been criticised, in our view successfully, in (Cartwright, 2017, sec. 5).

[2] Singular terms often occur as noun phrases, but school grammar is not useful for semantic analysis. The most obvious difference is that from a logical point of view, in expressions for relations, as in 'Carl is taller than Ann' both names are singular terms, but in school grammar are they treated as different grammatical parts, 'Carl' is the noun and 'Ann' is the direct object.

5.2.2 Generalised Causal Relations

In science one usually wants generalisable knowledge, one wants to be able to make inferences from particular phenomena to general states of affairs. In doing so one must organise data about individual cases using classifications of some sort. When making such classifications we use *variables*, categorical or quantitative. A number of individual cases belonging to the same category or quantity constitute the values of the chosen variable.

So for example, a differential equation of the form

$$\frac{dy(t)}{dt} = ky(t)$$

states a generality: for each time point t in a given interval, the value of the state variable $y(t)$ is proportional to its own time derivative. This statement has the logical form 'For all times t in an interval, the function $y(t)$ is proportional to the derivative $\frac{dy}{dt}$', which is a statement about pairs of properties attributable to some systems or perhaps only to one particular system. Even in the latter case it is a general statement, since it is a generalisation over a system's states during a certain time period.

By spelling out the full expression for a functional relation we directly see that it does not contain any causal information. But such equations are often unconsciously given a causal interpretation. Such a causal interpretation requires additional information.

Correlation reports are similarly general statements, because the correlation is a relation between two variables. And it is well known that a correlation by itself is not enough for inferring a cause-effect relation.

Sometimes such extra information is at hand, in which case we say that one variable is the cause of another variable. (This topic will be further discussed in Sects. 7.1–7.3, and 8.5.1) Thus we causally relate *abstract things*, i.e., *types* of events and states of affairs, universals, as they are called in philosophy. A trivial example is 'repeated exercise increases fitness'. The meaning of this is that a person can increase his/her fitness, i.e., he/she can cause increase of fitness, if he/she performs regular exercise. This statement is not about any particular exercising event, but about all instances of exercise; it is a general statement, and empirical research has given us solid evidence for this causal relation between two attributes of persons.

Some functional relations in science are called 'laws', 'regularities' or 'equations'. These are often interpreted as expressing causal relations. So for example, most people take for granted that Newton's second law, $f = ma$, says that the force f on a body with mass m is *the cause* of that body's acceleration a. This reading is however not mandatory and can in fact be strongly criticised. In any case, the validity of inferences using Newton's second law depends only on the fact that the equation $f = ma$ is an identity; whatever the letters f, m and a refer to, Newton's

law tells us that the left and the right hand side in any particular application of this formula are different expressions for the same number. Nothing in the equation says anything about causes and effects.

But we use laws when calculating what to do in order to achieve our goals. In such reasoning it is the action or intervention that is the cause of the desired outcome. More about this in Sect. 6.7.

If a causal conclusion is drawn from a particular equation, be it a law or not, it is based on tacit causal assumptions, not expressed by that equation. From a mere mathematical relation between values of variables, no causal conclusion can legitimately be drawn. Nancy Cartwright formulated this as 'No causes in, no causes out' (Cartwright, 1983).

5.3 Causation in Qualitative Studies

A few philosophers hold that one can, in some cases, directly observe a causal relation between two individual events. In Sect. 3.3 we have argued against that view, we hold that one cannot observe any relations at all; what we observe are objects and singular events, and in some cases we can tell the time ordering of a sequence of events/states of affairs. But this is obviously not sufficient for inferring any causal relations between such events/states of affairs; hence causal relations cannot be inferred *only* from singular case studies, irrespective of whether one collects quantitative or qualitative data.

Some researchers, such as Guba and Lincoln (1989) conclude that causation has no place in qualitative studies, whereas others, notably Joseph Maxwell (Maxwell, 2004, 2012, 2021) are of the opposite view.

The argument that causation has no place in qualitative research is based on two premises: (1) experiments with control groups are necessary for valid inferences to cause-effect relations and such experiments are not done in qualitative research, and (2) valid generalisations from singular case studies are not possible, unless background assumptions are invoked.

Data from a qualitative study concerns a very limited number of informants not being randomly selected from a population. So even if we obtain information about sequences of events from what the informants are saying, how do we know that the observations can be generalised? And even if one can, given some reasons for generalising the findings to other cases, how do we know that events are causally related? These sceptical reflections have led many qualitative researchers to hold that causation has no place in qualitative studies; qualitative studies aim at descriptions of individual phenomena only. However, when using ordinary language for these descriptions it is hard to completely avoid causal idiom. Maxwell writes:

> Becker described how it led qualitative researchers to use evasive circumlocutions for causal statements, "hinting at what we would like, but don't dare, to say" (Becker, 1986, 8). However, Hammersley argued that "in practice, virtually all qualitative researchers implicitly make causal claims, for example about what factors have 'influenced', 'shaped',

'formed', 'brought about', etc., some outcome" (Hammersley, 2012, 72). (Maxwell, 2021, 379)

Hammersley's observation is similar to the general observation we made in Sect. 2.2 that causal relations are indicated by quite a number of different expressions, e.g., 'bring about', 'produce', 'make happen', 'lead to' etc.

Successive cause-effect relations make up causal mechanisms, see Sect. 8.4. Those defending inferences to causal relations using qualitative data do so by referring to such causal mechanisms known in advance. The core idea is that a detailed description of the sequence of events in one or a few individual cases enables us not only to tell the time order of events, but also that we sometimes, using previous theory as background, can infer that they are causally related. Thus Miles and Huberman (1994, 147) writes:

> Qualitative analysis, with its close-up look, can identify *mechanisms*, going beyond sheer association. It is unrelentingly *local* and deals well with the *complex* network of events and processes in a situation. It can sort out the *temporal* dimension, showing clearly what preceded what, either through direct observation or *retrospection*.(emphasis in original).

Another researcher writes:

> Causal arguments are usually framed in terms of the effects of variables on each other. However, developmental, mechanical, processual, and comparative arguments all imply something about why and how social phenomena or processes occur or operate, and in this sense qualitative research does deal with questions of causality, although very often it wishes to think and speak of it in a different way. In fact, many have argued that qualitative research is particularly good at understanding causality, again precisely because of its attention to detail, complexity and contextuality, and because it does not expect to find a cause and an effect in any straightforward fashion. (Mason, 2018, 222)

The claim is thus that attention to 'detail, complexity and contextuality' may provide information about causal relations. It may do, but it requires background knowledge. Just as in quantitative studies, inferring a cause-effect relation requires more information than mere observations of one or a few individual cases. Cartwright (1983) concluded: 'No causes in, no causes out.'

5.4   Causation and Feedback Loops

Feedback loops might seem to contradict the condition that causes precede their effects. That is a mistaken conclusion.

An individual cause precedes its effect. In the limit, cause and effect may, for all practical purposes, be simultaneous, but it is not possible that the effect precedes its cause. The reason is simple: if we know, or have strong reason to believe, that two events or states of affairs are related as cause and effect, we label them 'cause' and 'effect' according to their timing; the first occurring is the cause of the other one, the effect. A component of the meaning of the word 'cause' is that it is followed in time by its effect. This is merely another way of stating the point made at the

beginning of this chapter, namely, that '....causes...' is an asymmetric relational term. This asymmetry is based on the fact that all transmission of signals takes time and causation requires signals.

The time order of cause and effect follows from two facts: (1) the cause must do something in order to bring about its effect, i.e., sending a signal of some sort, and (2) signals travel with at most the velocity of light.

In physics it is uncontroversial that a cause and its effect are connected by a signal carrying a conserved quantity, for example energy, being transmitted between cause and effect. There is an upper limit for the velocity of such signals, the velocity of light in vacuum; this is a consequence of relativity theory. This means that if two events occur at times and places such that no signal could have been transmitted between them, they cannot be causally related.[3]

Cause and effect must be connected by a physical signal also when we discuss causal relations in biology, psychology, sociology or history. It is for example often said that the cause of the First World War was the assassination of archduke Franz Ferdinand in Sarajevo on June 28, 1914. The physical signals, such as telegrams sent from Sarajevo to Vienna, Berlin and other power centres, are not salient in discussions about the causes of the First World War. But without any such transmission of information via physical links there would not be any causal link between the assassination and the outbreak of that war. So the physical link is a *necessary condition* for a causal relation between two events, but such links are often not *salient* in discussions about causes in history. The topic will be further discussed in Sect. 5.6.

The time order of cause and effect might appear to conflict with the idea of feedback loops, but that is not so; in fact feedback loops presuppose that each individual cause precedes its effect in time.

A feedback is usually described as that a cause A brings about its effect B, which in turn causes A. This is confusing. What is meant by a feedback loop should properly be described by talking about individual events. Event 1 causes Event 2, which in turn causes Event 3, which in turn causes Event 4, etc. This may be described as a feedback loop if for example Event 1, Event 3, Event 5, etc., are individual cases of the same *type* of events, let's say A, and if Event 2, Event 4, Event 6, etc., all belong to another *type* of events, B. A system may change from a particular state s_1 to another state s_2 at a certain time interval and that state change causes another system to change its state at a somewhat later time, and this in turn causes the first system to return to its initial state s_1. The process may continue in many loops.

The point is that saying that a system is in the same state at several different times, we are talking about the same *type* of state. Around ten o'clock every day

---

[3] There occur in quantum physics so called non-local correlations in which two seemingly distinct events are strictly correlated without there being any common cause and which can be proven to occur within a time interval too short for a signal to pass from one event to the other. This seems to contradict the idea that causation can be transmitted by at most the velocity of light. That is however a false conclusion, see e.g. (Johansson, 2021, ch. 14).

I want coffee; I am in the same *type* of state every day, but each day my state of wanting to drink coffee is distinct from all the other instances of the same state type; they occur at different times.

When talking about types of states and events we have no timing in the descriptions, since types of events and states are not individual things, they are abstract entities not occurring in space and time.

The statement that a type of events/state of affairs causes another type of events/state of affairs is to be interpreted as that each individual cause precede its particular effect in time. Hence, feedback loops do not contradict the idea that an individual cause precedes its effect in time. In fact, feedback loops *presupposes* just that.

5.5 Causation and Probability

We often talk about probable causes of events and there is a connection between conditional probability and causation. The basic idea is that if a type of events A causes another type of events B, then $p(B|A) > p(B)$, i.e., the probability of B conditional on A is higher than the unconditional, or so called *marginal probability*, $p(B)$. (But the converse need not be the case!)

When we attribute probabilities to singular events, the latter must be described in some way. One and the same event may be identified by different descriptions and this affects its probability. Here is an illustration.

According to the records from WHO, circa 1% of all those who has had Covid died of this disease. So one may say that the probability that a certain Covid-infected person N will die is circa 1%.

But suppose we know the age of N, he is 20 years old. We may then use data of Covid deaths by age groups.[4] According to these figures the risk of dying of Covid in that age group is much, much smaller; only 42 deaths of 22,600 belong to his age group.(These are figures from Sweden, but the general trend is general, young persons have a much lower risk of dying of Covid than older people.) In other words, the probability of N dying of Covid, given that he is infected and belongs to the age group 20–29 years of age is $42/22{,}600 \cdot 0,01 \approx 0.00002$.

It is obvious that the probability for a certain event crucially depends on how we classify that event, see e.g., (Hájek, 2007).[5] It seems reasonable to conclude that if we had a complete description of an event (and of the individual(s) involved), its probability would be either zero or one.

---

[4] https://www.statista.com/statistics/1107913/number-of-coronavirus-deaths-in-sweden-by-age-groups.

[5] This fact very strongly indicates that probabilities should not be thought of as attributes of events per se, but of events as described in specific ways.

This dependence on classification of events is called the *reference class problem*. It has been called a 'problem' since it is a problem for those who think that probabilities are objective properties of events per se. In our view, this is in most cases wrong; probability attributions depends on our knowledge.[6] It is no problem that probabilities of events depend on how we classify them.

So far the discussion about probabilities has been based on the frequency interpretation of probabilities. There are other meanings of probability, though. One alternative is probability as degree of belief. Consider for example a conversation about politics held at the end of 2022 where one person asks another how probable she thinks it is that the war in Ukraine will end before the end of 2023. Whatever the answer, the probability is naturally interpreted as a degree of belief in the mind of the respondent, not on any observed frequencies of the lengths of wars. Thus the probability assignment is to a person's mental state. Another person may have another degree of belief. Differences between different person's degrees of belief need not be based on different reference classes, they might be subjective estimates based on all sorts of information, or perhaps none at all.

Such subjective probabilities are however not very often used in scientific discourse; the great majority of talk about probabilities in science is based on relative frequencies. Briefly: the probability for a type of event A, is the relative frequency of individual events of type A in an infinite series of trials/tests/observations of this event type. Since we cannot perform an infinite number of observations we need a method to calculate probabilities from observed frequencies in observed samples. This is presented in the next chapter and in Appendix C.

To repeat, if A causes B, then $p(B|A) > p(B)$, if the probabilities are interpreted as relative frequencies in populations. But the converse is not true; from $p(B|A) > p(B)$ we cannot infer that A is a cause of B. The reason is that this inequality shows no more than that A and B are correlated, and a correlation can occur without there being any causal link between the correlated events; there might be a common cause, in medicine and other fields usually called a confounder. More about that in the next chapter.

Summarising, talk about the probability of an event in scientific contexts presupposes in most cases that event is classified as belonging to a reference class. This is so since the probability of an event is defined as the relative frequency of this *type* of events in the reference class.

5.6   Many Causes: INUS-Conditions

The statement 'A caused B' does not entail that A was the only cause of B. Almost any event, state of affairs or variable may have several causes. One way to bring some order among these is to investigate their time order. This may enable us to

---

[6] There is one exception, namely, transition probabilities in quantum theory, see (Railton, 1981).

discern causal chains, which is possible since causal relations are transitive; if A is a cause of B and B is a cause of C, then A is a cause of C. If need be, we explicitly say that A is an indirect cause of C.

Another way of bringing some order is to separate between causes and background conditions. This is a pragmatic distinction. Suppose we have good reason to say that all of $A_1, \ldots, A_n$, are causes to a certain event or state of affairs B. But what to do with this information? When we want to do something about B, to make an intervention, it is good to know what is the easiest, or most effective intervention to do. When such information is at hand and one has selected one cause, say A_k, as the cause (the most important cause, or the most effective cause etc.), one may, for all practical purposes, treat the other factors as background conditions.

A very illuminating illustration can be found in (Hesslow, 1984), see Fig. 5.1. We see that some fruit flies have shortened wings and when asking for the cause of this fact, the answer depends on which comparison one makes. If we take the temperature at 22 °C as a fixed background condition, one explains the shortened wings as being caused by the mutation. But if we compare the wing lengths at different temperatures we naturally say that the cause was the low temperature.

We may conclude that there are at least two necessary conditions for fruit flies to have shortened wings: (1) a mutation, and (2) being bred at room temperature. Which one of these conditions one selects as *the cause* depends on which comparison one makes. Hesslow's conclusion was that the distinction between genetic and environmental causes of diseases and other aberrant states depends on contrast, it is no real difference.

This example fits nicely into John Mackie's definition of cause as an INUS-condition (Mackie, 1965):

Def.   A CAUSE is an Insufficient but Necessary part of a complex of conditions, which together, as a complex, is Unnecessary but Sufficient for the effect.

One may observe the indefinite 'A cause'; it follows from the definition that a particular effect may have several causes all satisfying the definition.

Many, following Mackie, have distinguished between causes and background conditions, the latter also satisfying the definition given above, but not labelled 'causes' in specific contexts. It is rather obvious that the distinction between cause and background conditions is a pragmatic affair; one chooses one INUS-condition as the cause depending on ones particular interests, background assumptions or perceived contrast.

Illustration: The Discussion About the Causes of the Estonia Disaster
M.S. Estonia, a cruise-ferry built 1980, sailed on Estline's Tallinn–Stockholm route. The ship sank in stormy weather on 28 September 1994 in the Baltic Sea between Sweden, Finland and Estonia. It was one of the worst maritime disasters of the twentieth century, claiming 852 lives. A heated debate about the cause followed this disaster. A number of factors were mentioned:

1. The captain's decision to go full speed in the strong head-wind. (it was nearly full storm.)
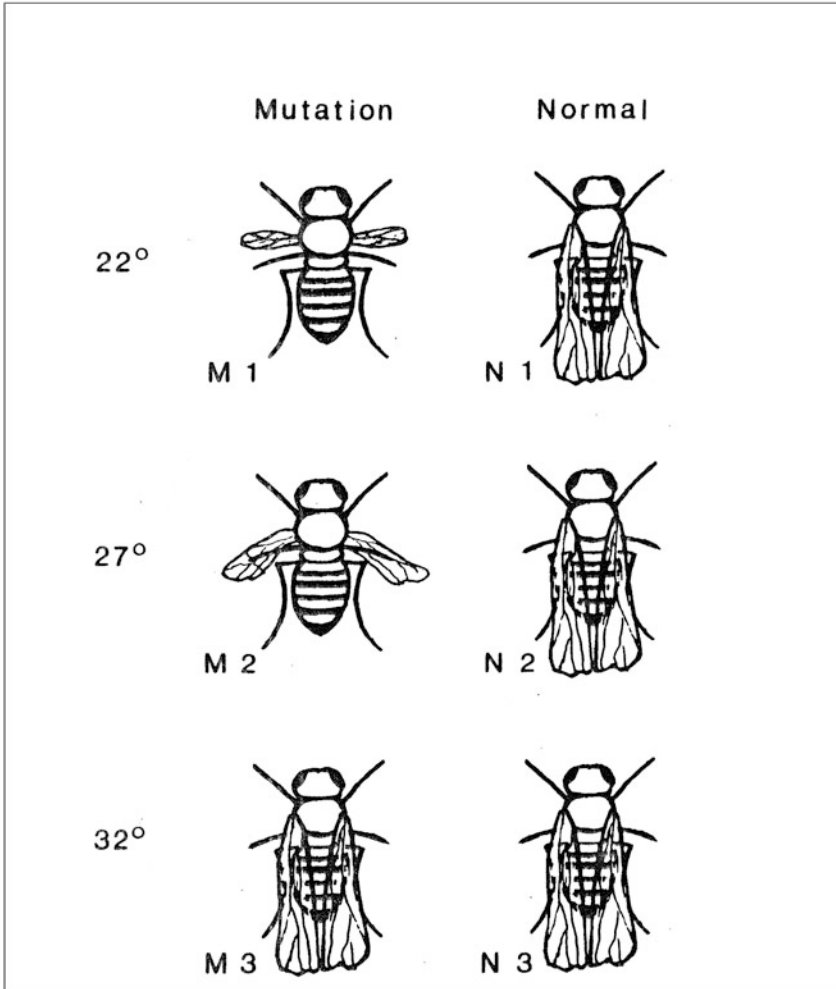
Fig. 5.1 Two breeds of Drosophila Melanogaster, one with normal genes and with a mutation. The two breeds have been grown in three different temperatures. Figure adapted from Hesslow (1984)

2. The construction of the bow was too weak. (Due to the pressure from the high waves it opened up.)
3. Estonia had an open car deck with no partitioning. This allowed all incoming water from the opened bow to flow to one side of the ship.
4. The ferry company's demand on the ferry captain to keep the time table, thus pressuring the captain to go full speed.

5. The decision made by the maritime classification society (Norske Veritas) to register this ship for traffic between Tallinn and Stockholm; its construction was not appropriate for this duty.

All these factors are clearly INUS-conditions for the disaster. An INUS-condition which never has been mentioned as the cause is the weather (!), the strong head-winds, producing waves of 10 m or more. Why was this condition never mentioned?

It is apparent that different agents chose different factors as being the main cause and that the different views depend on different perspectives and goals. Some agents wanted to pick who was legally and morally responsible; others were more interested in learning from this disaster and changing the construction of car ferries, the rules for car ferries, security prescriptions etc.

One cannot really blame the weather, nor do anything about it, so no one mentioned the weather as the cause. This is clearly a difference from the views of our ancestors; similar disasters, ships going down during storms and causing many deaths, have in most cases in history been explained by stormy weather.

The selection of 'the cause' or 'the main cause' among these factors is clearly made from an agency perspective. People want to know the cause, or the most important cause, or the salient cause, because they want to take action. Some relatives to the diseased wanted to know who is responsible in order to start a court trial, shipping authorities wanted to know what could be done in order to prevent similar catastrophes in the future, etc.

5.7 Summary

Causation is primarily a relation between individual events or states of affairs. Secondly it may obtain between types of events, states of affairs and variables. If the latter is the case, there must be causal relations between individual events/states of affairs making up these types and variable values.

A merely functional relation between two variables is not sufficient for concluding that these variables are causally related; one need more information in order to interpret an equation as being based on a causal relation.

If a type of events or state of affairs is a cause of another type of event or state of affairs, that cause increases the probability of the effect; prob(effect|cause) >prob(effect).

Most events have many INUS-conditions. In any particular case one or few of these are labelled 'cause', the rest is treated as background conditions. This distinction is based on pragmatic considerations.

Discussion Questions

1. Can you imagine a case where the distinction between cause and background conditions is independent of background beliefs, goals or interests on the part of the person who makes this distinction?
2. Could there be cases where people agree on what the cause of an undesirable state of affairs is, whereas the most efficient remedy is to do something about some factor among the background conditions?
3. That cause and effect must be physically connected is rather obvious when we talk about causal relations in nature. But what about historical, political or social causes; is it possible that there is no physical contact between for example two historical events, while still they could be related as cause and effect?

References

Becker HS (1986) Writing for social scientists. University of Chicago Press, Chicago
Cartwright N (1983) How the laws of physics lie. Oxford University Press, Oxford
Cartwright N (2017) Can structural equations explain how mechanisms explain? In: Beebee H, Hitchcock C, Price H (eds) Making a difference: essays on the philosophy of causation. Oxford University Press, Oxford, Chap. 8
Guba EG, Lincoln YS (1989) Fourth generation evaluation. Sage Publications, Newbury Park
Hájek A (2007) The reference class problem is your problem too. Synthese 156(3):563–585
Hammersley M (2012) Qualitative causal analysis: grounded theorizing and the qualitative survey. In: Cooper B, Glaesser J, Goom R, Hammersley M (eds) Challenging the qualitative-quantitative divide. Explorations in case-focused causal analysis. Continuum, London, pp 72–95
Hesslow G (1984) What is a genetic disease? On the relative importance of causes. Nordenfelt L, Lindahl BIB (eds) Health, disease and causal explanations in medicine. Reidel, Dordrecht, pp 183–193
Johansson LG (2021) Empiricism and philosophy of physics. Synthese library, no. 434. Springer Nature, Cham
Mackie J (1965) Causes and conditions. Am Philos Q 2(4):245–264
Mason J (2018) Qualitative researching, 3rd edn. Sage Publications, Los Angeles
Maxwell JA (2004) Using qualitative methods for causal explanation. Field Methods 16(3):243–264
Maxwell JA (2012) The importance of qualitative research for causal explanation in education. Qual Inquiry 18(8):655–661
Maxwell JA (2021) The importance of qualitative research for investigating causation. Qual Psychol 8(3):378–388
Menzies P (2012) The causal structure of mechanisms. Stud History Philos Mod Sci C 43(4):796–805
Miles MB, Huberman AM (1994) Qualitative data analysis: an expanded sourcebook. Sage Publications, Thousand Oaks
Railton P (1981) Probability, explanation, and information. Synthese 48(2):233–256

Chapter 6
Causation, Laws and Regularities

Abstract In this chapter we explore the connections between on the one hand causal relations and on the other hand strict and less strict laws, i.e., regularities, expressed as correlations and regressions.

It is tempting to think that laws and regularities describe general causal relations. They do not. Neither laws nor regularities distinguish between cause and effect, they state relations between quantities only; the causal aspect is connected to the manipulation and this aspect is not represented in formulations of laws and regularities.

Non-strict laws, often called 'regularities', differ from strict laws in that they are conditioned on ceteris paribus clauses, i.e., unspecified clauses of the form 'all else the same'. This makes generalisations, i.e., inferences to unobserved situations, difficult.

The main points of this chapter are:

- Laws, strict and non-strict, express relations between quantities, not cause-effect relations.
- Regularities are expressed by two measures, coefficient of correlation and regression.
- Inferring causal relations from laws and regularities requires additional information.
- Having such information one may represent causal relations using directed graphs and/or structural equations.

6.1 Laws and Causation

Many scientists and philosophers have assumed that general causal relations are expressed by causal laws. That is wrong, which was first realised by Bertrand Russell:

> 'The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm. (Russell, 1913).

It is not obvious what Russell meant by 'the law of causality', although it had been a standard phrase for a long time in philosophy. In any case, we take it that he held that there are no such things as causal laws in science. If so, he was basically right. Physics and chemistry do not contain anything that rightly could be called a 'causal law' and it is doubtful if there are any laws whatsoever outside physics and chemistry.

But we use laws (or weaker connections, *regularities*, to be discussed in Sect. 6.3) when making inferences about cause-effect relations. Physical and chemical laws connect variables to each other, or state invariance principles, and these laws are used in derivations.

The point is that several criteria must be satisfied for establishing a general cause-effect relation. The mathematical connection between the variables is only one of these conditions. A law, expressed as an equation, (or a regularity expressed as an equation including a random variable) is a *necessary* but not sufficient condition for there being a causal relation between the variables. In order to see this more clearly we may consider an episode in the history of science, Boyle's discovery of the law named after him.

Robert Boyle (1627–1691) studied the connection between pressure and volume of gases using a J-formed tube filled with mercury, as in Fig. 6.1. By adding mercury in the open end of the tube he changed the pressure on the air in the closed end of the tube. The volume of the air was measured by the scale on the left leg of the tube. He found that the product of pressure and volume is a constant, abbreviated as $pV = constant$, which is Boyle's law.

This simple mathematical relation between two quantities does not say which is the cause and which is the effect. This distinction is made first when identifying what is *manipulated* in a particular concrete case. In the experiments Boyle reported in (Boyle, 1662) that he *changed* the pressure by increasing the amount of mercury and passively *observed* the volume of the air; hence the pressure change is the cause and the volume change is the effect in this experiment. In another experiment it could be the converse.

It is obvious from the mere form of Boyle's law that it doesn't say anything about cause and effect. But the law is needed for establishing cause-effect relations involving changes of pressure and volume of gases.

This is a general trait of laws; mathematical relations between quantities only tell us that a change in any of the variables *logically entails* changes of at least one of the other variables. But mathematical-logical relations are not the same as causal relations.
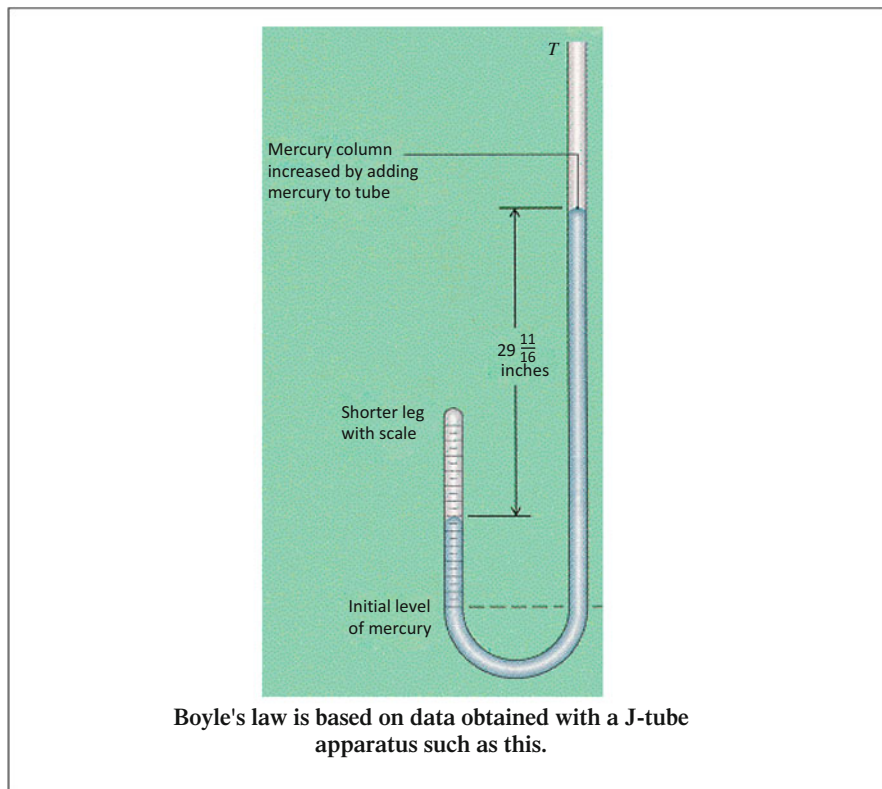
Boyle's law is based on data obtained with a J-tube apparatus such as this.

Fig. 6.1 Boyle's experiment, figure adapted from https://chemed.chem.purdue.edu/genchem/topicreview/bp/ch4/gaslaws3.html

6.2   Laws, Regularities and Ceteris Paribus Clauses

6.2.1   The Form of Laws

Boyle's law is one among a number of laws in physics and chemistry. These laws have the common feature of being general statements relating a number of quantities to each other, see (Johansson, 2019). But the generality is never explicit; usually only the numerical relations between quantities are explicit in law statements, while it is tacitly understood that these quantities are attributes of real objects of some kind. Here are some examples of physical laws expressing relations between quantitative variables:

1. Newton's second law:

$$f = ma \tag{6.1}$$

2. Coulomb's law:

$$f = k\frac{q_1 q_2}{r^2} \tag{6.2}$$

3. Maxwell's equations:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon} \tag{6.3}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{6.4}$$

$$\nabla \times \mathbf{E} = \frac{\partial \mathbf{B}}{\partial t} \tag{6.5}$$

$$\nabla \times \mathbf{B} = \frac{4\pi k}{c^2}\mathbf{J} + \frac{1}{c^2}\frac{\partial \mathbf{E}}{\partial t} \tag{6.6}$$

These equations, stating relations between quantities, are to be understood as abbreviations for full law statements. The quantities are in any particular application attributed to physical objects, i.e., bodies or fields. So the full verbal formulation of a physical law always contains a generalisation over all objects that can be attributed such quantities. So for example, Newton's second law is the following more complete statement:

Newton's second law: For any body with mass m, acceleration a and upon which a total force f acts, it holds that $f = ma$.

Thus, the logical form of scientific laws is that of *universally generalised conditionals*, UGCs for short. They do not tell us anything about causal relations; they merely inform us about numerical relations between some quantities attributed to a set of objects. They are general statements since they are true of all objects in a domain.

UGCs: Universally Generalised Conditionals
A conditional is a sentence of the form 'if A, then B', where A and B are complete sentences. (The word 'then' is often omitted.)
   Example: If it is raining, then the ground is wet.
   A universally generalised conditional is a sentence of the form
   'For all x, if x is A, then x is B'.
   Example: For all x: if x is a human, x has a heart.
   In logic, the expression 'For all' is called 'the universal quantifier'; hence a sentence of the form shown above may be called a universally generalised conditional.

A conditional is true if and only if either the antecedent is false or the consequent is true. The conditional doesn't say anything about what *makes* it true; whether it follows from mathematical or logical axioms, or if the meaning of B is included in the meaning of A, or whether A causes B. The same goes for UGCs.

One may now ask whether there are any tacit conditions for such general statements? In the case of Boyle's law, several researchers, the first was Amontons (1663–1705), discovered that the pressure of a gas depends on its temperature if the volume is constant. So a tacit assumption in Boyle's law is that temperature is constant. By combining these two relations and introducing the amount of matter, the number of moles, n, we arrive at the general law of gases, $pV = nRT$.

Is this the final truth about gases? No. In extreme conditions, for example at very high pressure or high temperature, one must take into account quantum effects, which entails some adjustments, resulting in van der Waals' law.

6.2.2 Strict and Not-So-Strict Laws

The process of adjustments and improvements of laws has in some cases come to an end, or so we believe. When all tacit conditions have been made explicit in the antecedent of a law, we have arrived at a strict law. So we distinguish between strict and not so strict laws, the former being those where we believe no further adjustments are needed.

But there are a vast number of not so strict connections between variables in the sciences. That they are not strict means that there are unknown but relevant conditions that have not been incorporated in the antecedent of the law statement. These unspecified and unknown conditions are sometimes referred to by a *ceteris paribus* clause. (This Latin expression means 'other things equal') The crucial thing is that we have not complete knowledge about such conditions; for if we knew them, we could incorporate all into the antecedent of a strict law, just as temperature was combined with Boyle's law, resulting in the general law of gases. So by saying that an observed regularity obeys a certain law, *ceteris paribus*, we indicate that we recognise the possibility of refinements, or even radical changes, in the so far not so strict law.

We think it better to call such non-strict laws with unknown scope of validity 'regularities' instead. This is less committing; calling something a regularity leaves open for changes and/or restriction of scope. Woodward (1997) suggests instead the label 'restricted invariances'.

Instead of referring to ceteris paribus conditions, one may add a random variable, an error term, to an equation expressing a not so strict relation between the variables.
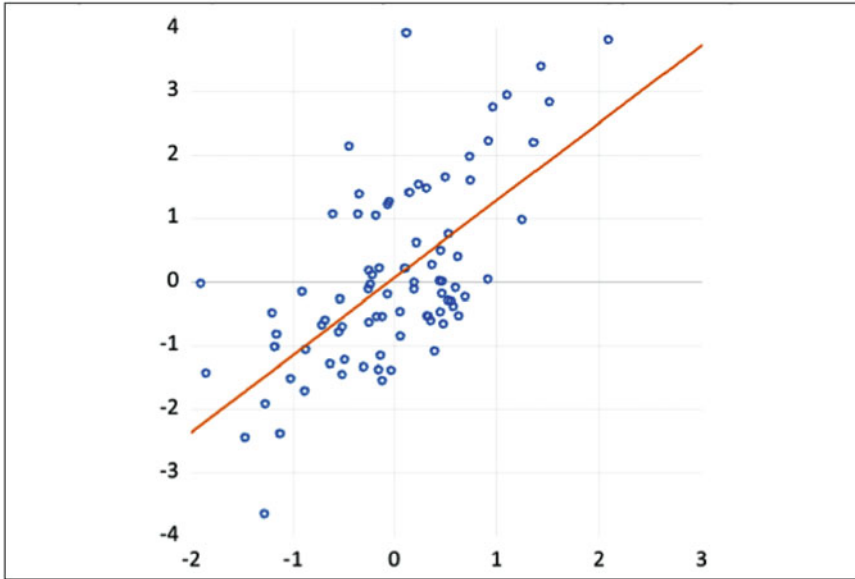
Fig. 6.2 Correlation between Indicator of Quality of Student Achievement (IQSA) (x-axis) and economic growth (EG) (y-axis), adapted from (Burhan et al., 2023)

(Talk about randomness is in most cases another way of saying that there are factors about which we at present lack information.)[1]

6.3   Correlation, Regression and Causation

Strict laws, in the sense of UGCs without any ceteris paribus clause or random variable, are so far not found in any discipline outside physics and chemistry. In e.g., biology, ecology, sociology and economics only weaker connections, regularities, have been found. In statistical terms such regularities are described by two functions, correlation and regression.

A scattergram displays vividly the information contained in the coefficient of correlation and the slope of the regression line, see Fig. 6.2.

---

[1] There is one exception, probabilities for state transitions in quantum theory, which are believed to be genuine and irreducible random events.

The coefficient of correlation tells us how strong the connection between the two variables are. If the correlation is -1 or $+1$, one can with certainty derive the value of variable Y from information about the value of variable X, or vice versa. If the correlation is zero there is no connection at all. In Fig. 6.2, where the correlation is rather strong ($R = 0.74$) one can, for a chosen value of X, determine an interval for the corresponding Y, or vice versa. One may further conclude that there must be other variables connected to the development index, although they contribute less than how many years girls go to school to human development index.

The coefficient of correlation is a measure of the spread of data-points around the regression line. If all data points were on the regression line, the correlation would be 1 (or -1, if the slope is negative). If the data points are completely randomly spread over the entire area of the scattergram the correlation is 0.

This means that if we want to formulate this as a regularity, we should write something like

$$EG = constant \cdot IQSA + U, \qquad (6.7)$$

where U is a probability distribution function representing all other factors, known or unknown. It is obvious that if the random variation in U is big, the equation is not of much use. A well-known example from economics may be useful as further illustration.

A.W. Phillips published 1958 a well-known paper, (Phillips, 1958), which showed that inflation and unemployment are roughly inversely proportional. This result is called the 'Phillips curve', see Fig. 6.3, which is our own drawing.
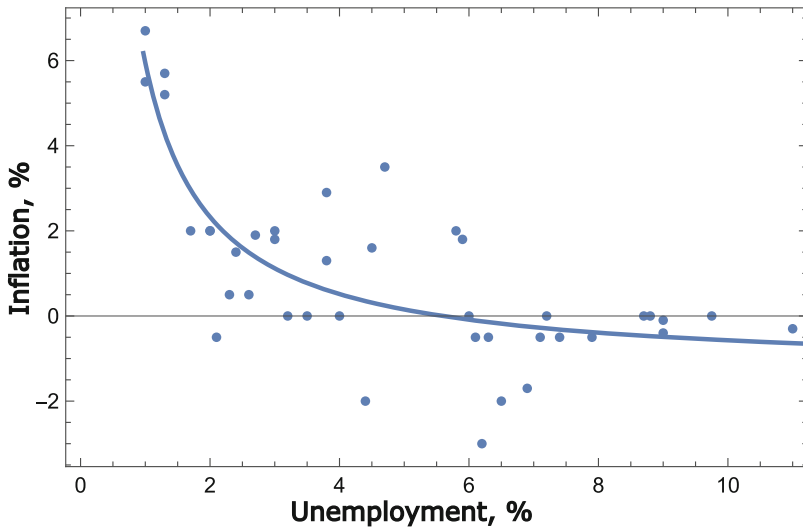

Fig. 6.3 Inflation vs unemployment in the UK 1861–1913

Paul Samuelson integrated this result into economic theory, see Samuelson (1983). For some time economic researchers thought that this relation was close to a real economic law, tacitly assuming that inflation can be manipulated in order to decrease unemployment. Policy makers throughout the western world used this result for policy decisions; when governments wanted to decrease the unemployment rate they increased expenditure and budget deficits, calculating that this would increase inflation and decrease unemployment. However, after some years it was realised that it didn't work as expected, often one got higher inflation without any decrease in unemployment. The conclusion was that the roughly inverse correlation had not been stable, hence some unknown ceteris paribus factor had changed. Here is quote from a review report published by FED:[2]

> Federal Reserve Chair Jerome Powell has been asked about the Phillips curve, during his July 2019 testimony before Congress. He noted that the connection between economic slack and inflation was strong 50 years ago. However, he said that it has become "weaker and weaker and weaker to the point where it's a faint heartbeat that you can hear now."
> In discussing why this weakening had occurred, he said, "One reason is just that inflation expectations are so settled, and that's what we think drives inflation." (Engemann, 2020)

Two things are pretty clear. The first is that since the data points are dispersed around the curve, there must be more factors than unemployment that determines inflation. This means that there is at most a probabilistic relation between unemployment and inflation. The second is that, since the connection has weakened over the years due to reduced inflation expectations, this factor, inflation expectations, was one of the unknown factors in the original study. Powell suggests that is the main cause of inflation (he used the word 'drives'). Hence, the Philips curve cannot be used as basis for political measures; it does not reflect a useful *causal* relation between high unemployment and low inflation.

A further obvious conclusion can be drawn: mere observational data, statistics, are not sufficient for inferring causal relations; one needs also other kinds of information. And since manipulability, or more generally, intervention, is strongly connected to causation, we need information from experiments, carefully designed interventions or natural experiments, in order to determine whether an observed correlation is a sign of a causal relation or not. The tools needed for such inferences are discussed in some detail in (Pearl, 2009). Here is a quote from this paper:

> Remarkably, although much of the conceptual framework and algorithmic tools needed for tackling such problems are now well established, they are hardly known to researchers who could put them into practical use. The main reason is educational. Solving causal problems systematically requires certain extensions in the standard mathematical language of statistics, and these extensions are not generally emphasised in the mainstream literature and education. As a result, large segments of the statistical research community find it hard to appreciate and benefit from the many results that causal analysis has produced in the past two decades. These results rest on contemporary advances in four areas:
>
> 1. Counterfactual analysis
> 2. Nonparametric structural equations

---

[2] Federal Reserve System, the central bank of United States.

3. Graphical models
4. Symbiosis between counterfactual and graphical methods.

(op.cit. pp. 97–98)

We have discussed counterfactual analysis in Chap. 4 (and suggested replacing potential outcomes for counterfactuals) and will bring up structural equations and graphical models in this chapter. Symbiosis of counterfactual and graphical methods will not be discussed in this book.

6.4   Correlations Between Boolean Variables

Boolean variables (after George Boole, 1815–1864) have only two possible values, such as true–false, yes–no, or 0–1. Boolean variables are common in social sciences, they are used when organising data in two categories (male–female, college education–no-college education, etc.). The measure of correlation between two Boolean variables is the ϕ coefficient ('mean square contingency coefficient').

Suppose we have two variables, X and Y, and denote their values '0' and '1' respectively. If we have n observations, we display the distribution as follows:

	y = 0	y = 1	Total
x = 0	n_{00}	n_{01}	$n_{0\bullet}$
x = 1	n_{10}	n_{11}	$n_{1\bullet}$
Total	$n_{\bullet 0}$	$n_{\bullet 1}$	n

It is rather obvious that if n_{00} and n_{11} together make up all the observations, so that n_{01} and n_{10} both are zero, we have a perfect correlation between X and Y. Likewise if the situation is completely reversed, all observations belonging to n_{01} or n_{10}. Thus, the ϕ coefficient is defined as:

$$\phi = \frac{n_{11}n_{00} - n_{10}n_{01}}{\sqrt{n_{1\bullet}n_{0\bullet}n_{\bullet 0}n_{\bullet 1}}} \tag{6.8}$$

So, just as with the usual coefficient of correlation ρ, the value is between -1 and 1, where the zero value means no correlation at all.

6.5   Directed Graphs and Structural Equations

The details of causal mechanisms may usefully be described using *directed graphs* and *structural equations*. Directed graphs visualise mechanisms and by using structural equations we can state quantitative relations between variables, i.e., we can give a measure of the strength of different causal connections.

6.5.1   Directed Graphs

Directed graphs is a conceptual and visual tool for displaying causal relations between, basically, values of variables. If we for example know that Z has two causes, X and Y, i.e., that the values of the variable Z is causally determined by the values of the variables X and Y, but not the other way round, we can visualise that with a directed graph of the form shown in Fig. 6.4. Figure 6.5 illustrates a situation where the variable Y has only one cause, the variable X, which in turn has only one cause, the intervention variable I. Directed graphs can be used to display rather complex structures, as is shown by e.g., Pearl (2000). Figure 6.6 is an example from his book (p. 215).

Assuming a free market economy, we can see that according to this model there are two ways to affect the price of a product: either to manipulate the wage costs for producing the product, or to manipulate the household income. If for example the household income is roughly the same during a certain period and the price have decreased, we may infer that it was caused by a decrease in wage costs. As always, we infer a causal relation between two individual events using information about causal relations between variables.
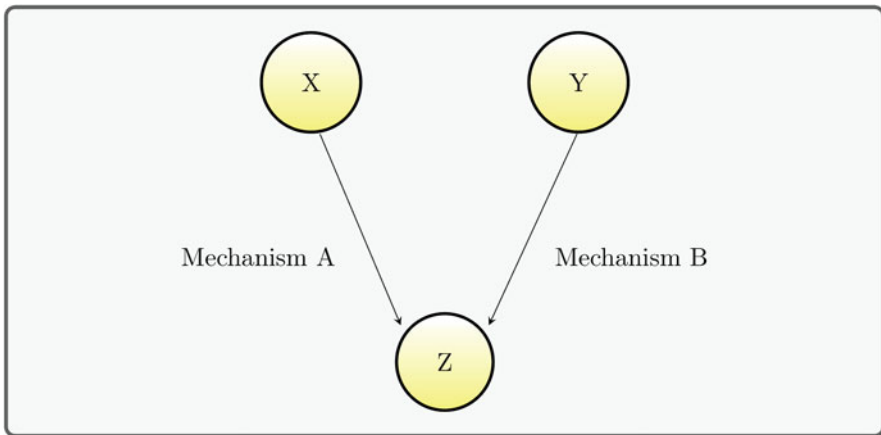

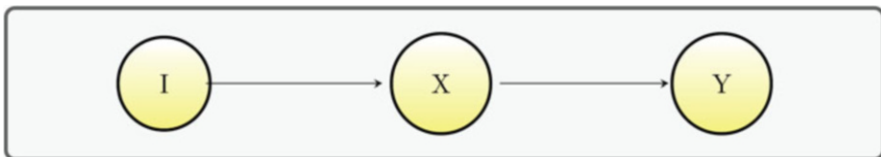**Fig. 6.4**   The variables X and Y are each individually contributing causes of Z

Fig. 6.5   The variable Y is directly caused by the variable X only, and the intervention I is the only direct cause of X. It means that the only way to change the value of X, and hence the value of Y, is to do something with I
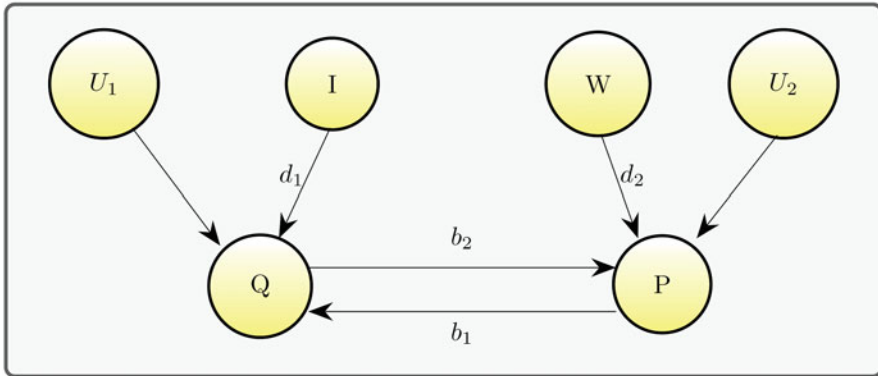
Fig. 6.6 A diagram depicting the causal relations between price (P) and demand (Q) for a certain product. U_1 and U_2 are unknown external factors, I the household income and W the wage costs for producing the product, see (Pearl 2000, 215)

6.5.2 Structural Equations

Let us not forget that a causal relation between two variables is based, in an ontological sense, on causal relations between individual values of these variables. Thus, the fact that the variable X is the only cause of Y means that the event that X has a certain value, say x_i, is the cause of the event of Y having the corresponding value y_i. From an epistemological point of view we go in the opposite direction: we first establish knowledge about causal relations between variables by performing experiments, which then enables us to infer a causal relation between a pair of individual events or states of affairs.

These relations can more precisely be represented by so called *structural equations*. The following equation represents the situation depicted in Fig. 6.4 (k_1 and k_2 are parameters giving the relative contributions from X and Y):

$$Z = k_1 X + k_2 Y \qquad (6.9)$$

Using linear equations is no substantial restriction. If the relation between an observed cause X and an effect Z is non-linear, one can easily make a variable transformation $X \rightarrow X' : X' = X + a_1 X^2 + ...a_n X^n$, so that Z is linearly dependent on X'. (All continuous functions, whatever their shape, can in any limited domain be approximated by functions of this type.)

Such equations differ from ordinary equations used in mathematical expositions of physics, economics and other 'hard' sciences in that the transformation rules of algebra are not valid in structural equations. The rule of interpretation for structural equations is that the left hand side represents the effect and the right hand side represent the cause or causes of this effect. This means that one cannot rewrite the

equation by moving terms from left to right hand side of '=', or vice versa, as is legitimate when using ordinary equations in derivations.

Therefore, using the identity sign, '=', in structural equations is not appropriate; it would be a better idea, and in fact necessary, to use an asymmetric sign, for example '=:' instead,[3] see (Pearl, 2000, 138) thus writing the equation above as:

$$Z =: k_1 X + k_2 Y \qquad (6.10)$$

Here the sign '=:' is to be read as '...is caused by ....', and the entire equation means 'The values of the variable Z are caused by the values of X and Y according to the weight factors k_1 and k_2'. A situation depicted as in Fig. 6.5 can be given as a system of equations:

$$\begin{cases} Y =: k_1 X \\ X =: k_2 I \end{cases}$$

and the relations depicted in Fig. 6.6 are given as

$$\begin{cases} Q =: b_1 P + d_1 I + U_1 \\ P =: b_2 Q + d_2 W + U_2 \end{cases}$$

(So there is a feedback mechanism here, see Sect. 5.4 and the discussion in Sect. 8.5.2.). It is obvious how to extend this to more factors and more steps.

Each step in such a chain of causal relations may be realised by different kinds of links. Such a chain of causes is a *causal mechanism*, and providing the mechanism connecting a cause and its final effect is a common way to respond to quests for a causal explanation, to be further discussed in Chap. 8.

6.5.3   Bayesian Networks

In Fig. 6.6 the two arrows between price (P) and demand (Q) go in opposite directions. This represents a mutual dependency between these variables, see the last equation system in the previous subsection. Is this mutual dependency due to causal mechanisms or not? In economics it is assumed, we believe, that there is a feedback loop here, meaning that the value of e.g., the variable Q at a certain time t_1 is causally dependent on the value of P at an earlier time, and a P-value at some time t_2 depends on earlier Q-values.

---

[3] Note the difference from ':=', which means 'is given the value' as used in some computer languages.
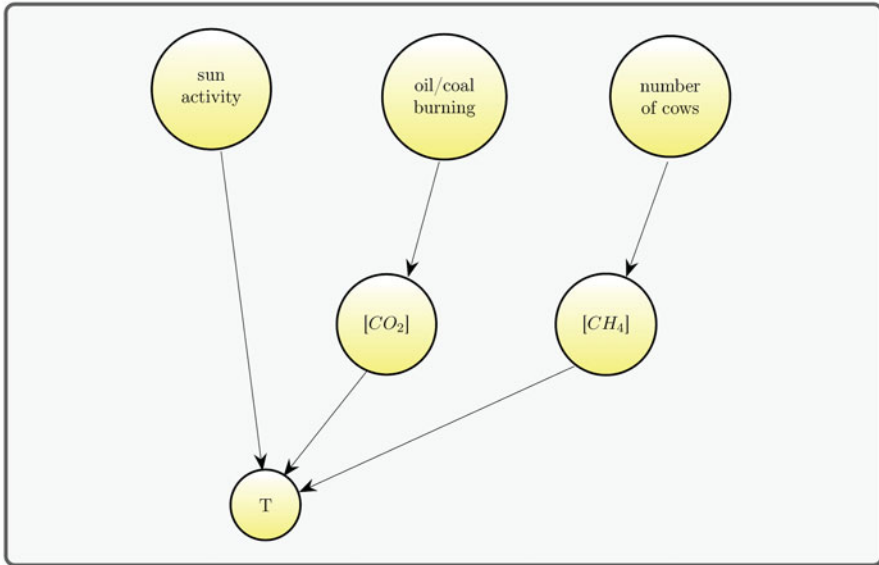
Fig. 6.7 A DAG depicting causal links from sun's activity, oil/coal burning and number of cows to the temperature in the atmosphere, T, via carbon dioxide and methane concentration. Observe that the three causal links are depicted as independent of each other, i.e., that the modularity condition is satisfied

If one knows, or has good reason to believe, that there are no feedback loops in a system, one may use Bayesian Networks, see e.g., (Pearl, 2000, Sect. 1.2), for modelling causal relations.

A Bayesian Network has two components, a *Directed Acyclic Graf* (DAG for short), and a set of conditional probabilities, one for each arrow in the graph. Figure 6.7 is a DAG and since there are five arrows, each representing a conditional probability for the connected variables, one needs information about five conditional probability distributions. For example, the left-most arrow connecting the sun activity and earth's temperature represents a conditional probability of the form $prob(T = t_1 | S = s_1) = p$, where T is the earth's temperature, S is the sun activity (in some measure) and p is the probability.

One should keep in mind Cartwright's 'no causes in, no causes out', (Cartwright, 1989). In other words, without causal assumptions as input in the construction of the network, one cannot draw any conclusions about causal relations from the network itself; it merely depicts statistical relations. (See further discussions about statistics and causation in Sect. 7.1.) But with input about causal relations, Bayesian Networks are useful tools for understanding causal structures and for making calculations.

When drawing the DAG one should ask oneself whether there are any causal interferences between different causal chains. In Fig. 6.7 there is no arrow between the concentration of CO_2 and of CH_4. The fact that no such arrow is drawn is a visualisation of the input, assumed to be correct, that there is no causal link between them. So when constructing the DAG, one needs to know whether there is any such link. The lack of causal couplings between different causal chains is in the literature called *modularity*, which is defined as follows:

> **Modularity**: If X_i does not cause X_j, then the probability distribution of X_j is unchanged when there is an intervention with respect to X_i.

This is related to the **Causal Markov Condition**, **CM**: (V is the set of variables in a Bayesian Network):

> **CM**: For all $X_i, X_j, i \neq j$ in V, if X_i does not cause X_j, then X_i and X_j are probabilistically independent conditional on the set of parents, pa_i, of X_i.

These two conditions are related, since given a set of extra assumptions one can derive CM from Modularity, see (Hausman and Woodward, 2004). Thus it is possible to perform a statistical test for the assumption that there is no causal link from X_i to X_j.

It may be observed that in constructing the figure we have taken for granted some causal relations, e.g. that cows produce great quantities of methane and this gas increases the temperature on Earth.

Why then is this network called 'Bayesian'? Because we use Bayes' theorem for updating the probabilities when new information is available.

(Barbrook-Johnson and Penn, 2022) contains a useful description of Bayesian Networks (in that book called 'Bayesian Belief Networks'). It contains a list of a number of softwares that may be used in constructing Bayesian Networks.

The authors rightly stress that the conditional probabilities connecting the nodes in the Network must be based on causal information, not mere observed statistics. Such information typically comes from stakeholders and they stress their importance:

> We must encourage users to acknowledge that BBNs are always dependent on stakeholder opinion (unless developed based solely on data) and that removing outputs from that context, and not making clear either the process, or the network (i.e. the model), from which they are derived almost always dooms us to see them misinterpreted. Even in cases where outputs are not misused or misunderstood, the appeal of the diagram of a BBN with conditional probabilities annotated can also lead many to view BBN and its associated analysis as a product, rather than a process. Not recognising the value in the process of using this method is to ignore at best half its value, at worst, all its value. (op.cit., p. 107)

6.6   Non-linear Dynamics

When studying the association[4] between two variables, the first step is to see how good the data points fit a linear regression line of the format $Y = a + bX$. The calculation of such a linear regression can be done using any statistical package.

When looking at a graph one sometimes gets the impression that a non-linear equation would fit better to the data points. (A more reliable method is to use a statistical package by which one can calculate the best fit of the data points to different functions.) So one may repeat the procedure with e.g., an equation of the form $Y = a + bX + cX^2$, or, as was the case with the Phillips curve, a function of the form $Y = a + bX^{-1}$. And one can go further and use non-linear equations of higher and higher degrees as mathematical models of the observations. (Feedback loops, see Sect. 5.4, is one mechanism that may generate non-linear dynamics)

So far, all equations discussed are continuous and one may think that discontinuous changes must be possible in the real world. Well, the question is not whether there really are discontinuous changes in reality, but whether there are so fast state changes that a discontinuous function is a good representation. If for example one can measure a dependent variable at most once a day, it may one day change so abruptly that a step function is a good description of its state evolution. One might assume that the variable had intermediate values in between the two measurements, but for predictive purposes it doesn't matter.

It is important to keep in mind that even if observational data quite well fit a non-linear equation, this fact in itself does not allow us to infer that the independent variable is a cause of the dependent variable. Just as is the case when a linear equation is a god fit, there may be common causes that produces the mathematical relation. The Phillips curve is a fine illustration; it is a non-linear equation, but, as we saw, there is virtually no causal connection between inflation and unemployment.

6.6.1   Predictions and Non-linear Dynamics

Non-linear evolution often surprises us, because we have a natural tendency to begin with the simplest hypothesis, a linear function, when investigating the relation between two variables. Consider, as an example, a simple physical experiment often made in physics courses in secondary school. The pupils are given a resistor, a current source, a current meter and a voltage meter. They are instructed to determine

---

[4] The term 'association' refers to purely statistical connections, correlation and regression. It is often mistakenly interpreted as a term for a causal relation.
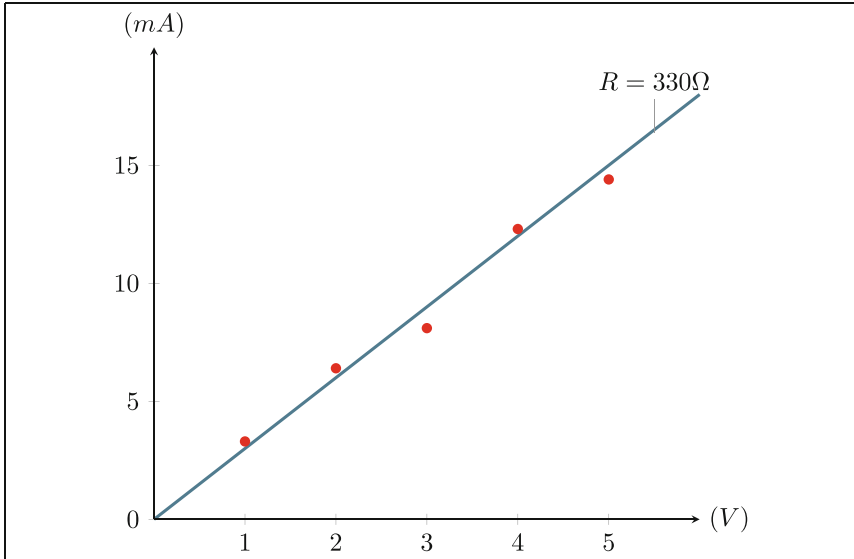
Fig. 6.8 Measurements of voltage and current in a resistor

the resistance of the resistor by making series of measurements of voltage and current. A typical outcome could be something like this:

Voltage (V)	Current (mA)
1.0	3.3
2.0	6.4
3.0	8.1
4.0	12.3
5.0	14.4

A graph of these results strongly suggests that the current is a linear function of voltage (Fig. 6.8).

In other words, one feels justified to conclude that the resistor has a constant resistance (R = U/I) of $R \approx 330\,\Omega$. However, if one continues measuring voltage and current, this inference may be proven wrong. For one common type of resistors one would get something like the following data:

Voltage (V)	Current (mA)
1.0	3.3
2.0	6.4
3.0	8.0
4.0	12.3
5.0	14.4
6.0	16.5
7.0	19
8.0	21
9.0	23
10.0	24
11.0	24.5

We see that the resistance increases as the voltage increases. Higher voltage leads to higher currents which results in warming, which leads to higher resistance. It is well known both from experiments and theory that the power expenditure in many materials, as measured by warming, is proportional to $current^2$, hence there is no linear relation between current and voltage, except as an approximation at low voltage (Fig. 6.9).

This is just a very simple example of a non-linear response where one can calculate a non-linear equation with good fit to any number of experimentally obtained data points. If this non-linear but continuous function can be guessed or derived from theory, one still has an explanation and can make good predictions.
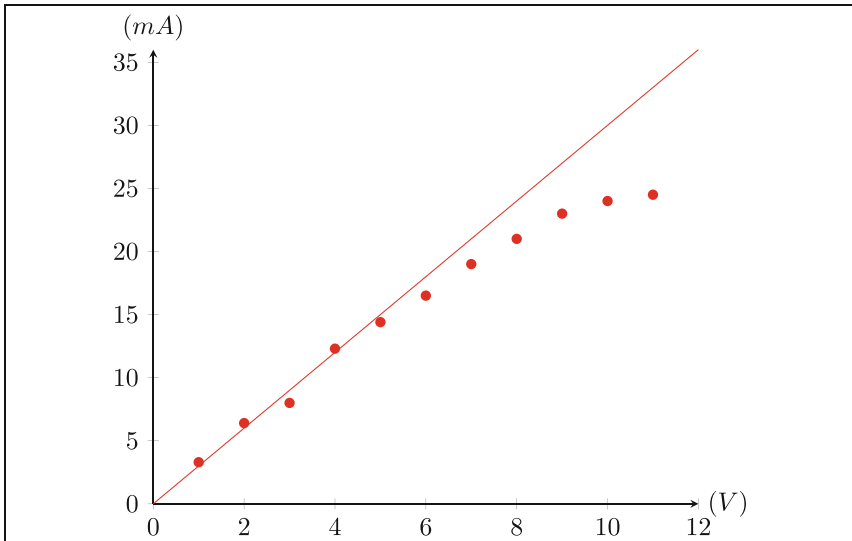

Fig. 6.9 Extended measurements of voltage and current in a resistor

In complex social-ecological systems there are many mechanisms, most of which are being poorly understood. There is seldom a possibility to perform controlled experiments and the data points are sparse. Having a small sample of data points which, just as in this simple example, suggests a linear relation between two variables, one naturally extrapolates this linearity to non-observed situations. But the extrapolation might prove wrong, the relation between the variables were not linear. In short, failed predictions is very often attributed to a non-linear and unforeseen connection between the predictor and the response variable.

6.7   Causation, Manipulation and Intervention

Our use of causal notions is basically connected to our interest in performing beneficial actions: we want to improve our conditions in all possible ways. Thus, several philosophers have suggested to define causation in terms of *manipulation*. The cause-effect relation is the relation between an action and the result of that outcome. Critics have objected that this is circular, 'manipulation' also expresses a causal notion. (Menzies and Price, 1993) countered this argument by pointing out that we have direct experiences of ourselves acting as agents:

> The basic premise is that from an early age, we all have direct experience of acting as agents. That is, we have direct experience not merely of the Humean succession of events in the external world, but of a very special class of such successions: those in which the earlier event is an action of our own, performed in circumstances in which we both desire the later event, and believe that it is more probable, given the act in question, than it would be otherwise. To put it more simply, we all have direct personal experience of doing one thing and thence achieving another. We might say that the notion of causation thus arises, not as Hume has it, from our experience of mere *succession*; but rather from our experience of *success*; success in the ordinary business of achieving our ends by acting one way rather than another. (Menzies and Price, 1993, 194)

The point is that the meaning of the term 'cause' and its synonyms is determined by being used in direct linguistic interactions between people in concrete circumstances.[5] This is true not only of 'manipulation', but also of many other terms with a clear causal sense, as thoroughly discussed in Chaps. 2 and 3. So we hold Menzies & Price' defence valid and it fits nicely with our observations in those chapters.

But, as Pearl observed (see the quotation in Chap. 2), we have since long extended our use of causal notions to cover also phenomena not in the scope of any possible human action. For example, the tides are caused by the motions of the sun and the moon, but, certainly, we cannot manipulate the motions of these celestial objects. This example indicates that we have generalised the concept of cause from covering merely human manipulations and their effects to a broader class of events. What, then, is the implicit idea behind this particular generalisation?

---

[5] This is a central point in (Wittgenstein et al., 1969). His slogan was 'Meaning is use'.

Extending the scope of a certain concept is always based on perceived similarities between old and new cases. In the case of extending the causal relation to be applicable to the connection between the motions of the moon, the sun and the tides, it is the physical links that is the basis.

The starting point is the application of the cause-effect relation to collisions between two bodies. The hit is the cause and the change of motion of the second body is the effect. Such events function as paradigmatic examples of cause-effect relations after the scientific revolution.

After Newton's *Principia* we further learnt that physical interactions may obtain at distance, transmitted by gravitational, electromagnetic and other fields. So when gravitation theory could be used to derive the tides, using the motions of the sun and the moon as input, this interaction was naturally classified as an instance of causation.

The Causal Link Between the Tides and the Motions of the Sun and the Moon

Tables of the tides in English ports were published already in 1555[a] if not earlier. The tables were calculated from the motions of the sun and the moon, which had been predictable since long, and the correlations between the tides and the positions of the moon and the sun were known. But it was not known how the motion of these celestial bodies could cause the tides. Explaining this was one of Newton's achievements. In his *Principia* (published 1687) he showed that by using the law of gravitation, applied to the water in the seas, the moon and the sun, he could explain the tides. In other words, he showed that there is a physical link, a force connecting these celestial bodies and the water in the seas. That was obviously sufficient for classifying this interaction as a cause-effect relation.[b] This is an example of how forces in general were conceived as mediators of causal effects.[c]

---

[a] https://www.bl.uk/onlinegallery/onlineex/unvbrit/t/001roy000017a02u00011000.html.

[b] Newton also correctly claimed that the tides also have an effect on the motions of the moon. But it is very small and can in almost all calculations be neglected. So the gravitational interaction is in this case in practice treated as an asymmetric relation and hence fitting the asymmetric cause-effect notion.

[c] Gravitation is in modern physics not conceived of as a force but rather an effect of the curvature of spacetime. There is no room for external interventions in general relativity, all matter and energy is included. That does not conflict with talk about causes when a more 'local' perspective is applied.

We have already discussed another important extension of the idea of causation as manipulation, namely, so called natural experiments. In such experiments no intentional manipulation of a variable is made. Some authors have thus introduced the term 'intervention' as a substitute for, or rather extension of, 'manipulation',

when talking about natural experiments. An intervention is a change of a variable that need not be the result of a intentional action by an agent.

Those who first introduced the concept of intervention had a rather restricted notion in mind, but it was soon extended. Here is Woodward's description of this evolution:

> Another important extension of interventionist ideas, also with a focus on inference but containing conceptual innovations as well, is due to Eberhardt (2007) and Eberhardt and Scheines (2007).
>
> These authors generalise the notion of intervention in two ways. First, they consider interventions that do not deterministically fix the value of variable(s) intervened on but rather merely impose a probability distribution on those variables. Second, they explore the use of what have come to be called "soft" interventions. These are interventions that unlike the fully surgical ("hard") interventions considered above (both Pearl's setting interventions and the notion associated with M1–M4), do not completely break the previously existing relationships between the variable X intervened on and its causes C, but rather supply an exogenous source I of variation to X that leaves its relations to C intact but where I is uncorrelated with C.
>
> Certain experiments are naturally modelled in this way. For example, in an experiment in which subjects are randomly given various amounts of additional income (besides whatever income they have from other sources) this additional income functions as a soft, rather than a hard intervention. Soft interventions may be possible in practice or in principle in certain situations in which hard interventions are not. Eberhardt (2007) and Eberhardt and Scheines (2007) explore what can be learned from various combinations of soft and hard, indeterministic and deterministic interventions together with non-experimental data in various contexts. Unsurprisingly each kind of intervention and associated data have both advantages and limitations from the point of view of inference. (Woodward, 2016)

The list M1–M4 of conditions for interventions referred to above is as follows (I = Intervention):

(M 1) I must be the only cause of X; i.e., the intervention must completely disrupt the causal relationship between X and its previous causes so that the value of X is set entirely by I.

(M 2) I must not directly cause Y via a route that does not go through X.

(M 3) I should not itself be caused by any cause that affects Y via a route that does not go through X.

(M 4) I leaves the values taken by any causes of Y except those that are on the directed path from I to X to Y (should this exist) unchanged.

Drawing conclusions about causal relations from statistical information is a central task in much of empirical science and several books have been written about this topic. Some useful ones are (Freedman et al., 2010), (Hernan and Robins, 2020), (Imbens and Rubin, 2015), and (Illari et al., 2011).

6.8   Summary

A common view is that scientific laws express causal relations. This is wrong; scientific laws state numerical relations between quantities such as mass, energy, momentum etc., but they do not express any causal relations between these quantities. The distinction between cause and effect is based on which quantity we directly manipulate in any concrete situation. The change of a variable performed by a certain manipulation is the cause and the change of some other variable, which according to a scientific law must change, is the effect. It follows that from a merely observed regression or correlation one cannot infer any causal relation. We need further information, telling us which interventions have been made, in order to draw any valid conclusion about a cause-effect relation.

Thus, mere statistical information, i.e. knowledge about correlations and regressions between two variables, is not sufficient evidence for a causal relation between them.

The argument applies also to non-quantitative variables. One can calculate statistical measures, such as Chi-square numbers, applicable to Boolean variables or variables over rank ordered data and estimate statistical dependencies. But the general lesson applies: statistical dependencies are insufficient for conclusions about causal relations.

The available data may allow a generalisation in the form of an ordinary equation where one variable is a function of one or several others, but this is not sufficient for taking the independent variables in that equation to be the causes of the dependent variable. The basic reason is that an ordinary equation, i.e., an identity sign flanked by two mathematical expressions, is symmetric; there is no asymmetry in the identity sign '='. But structural equations aim to distinguish between left and right; the left hand side is thought to represent the effect and r.h.s the total cause. Thus structural equations differ sharply from ordinary equations, which is a strong reason not to use the common identity sign '=' in structural equations.

A structural equation represents the asymmetry of cause and effect, and this asymmetry is postulated, hypothesised or empirically proven as the *reason* for formulating the structural equation.

Discussion Questions

1. Why are experiments considered necessary for reliable inference to causal relations?
2. Are there any scientific laws that clearly and explicitly express a causal relation?
3. Are there any scientific laws that has *not* the form of a universally generalised conditional?
4. Quite often one observes two correlated variables. What is required for us to infer that this correlation is due to a causal relation between them?

5. Similarly with regression: If $Y = aX + b$, under what conditions can one infer that X is a cause of Y?
6. What is the difference between manipulation and intervention?

References

Barbrook-Johnson P, Penn A (2022) Systems Mapping: How to build and use causal models of systems. Palgrave Macmillan, London

Boyle R (1662) New experiments physico-mechanical, touching the spring of air, and its effects. Printed by H. Hall for Tho. Robinson, Oxford

Burhan NAS, Sabri MF, Rindermann H (2023) Cognitive ability and economic growth: how much happiness is optimal? Int Rev Econ 70:63

Cartwright N (1989) Nature's capacities and their measurement. Clarendon Press, Oxford

Eberhardt F (2007) Causation and intervention. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA

Eberhardt F, Scheines R (2007) Interventions and causal inference. Philos Sci 74(5):981–995

Engemann K (2020) What is the Phillips curve (and why has it flattened)? https://www.stlouisfed.org/open-vault/2020/january/what-is-phillips-curve-why-flattened

Freedman DA, Collier D, Sekhon JS, Stark PB (2010) Statistical models and causal inference. A dialogue with the social sciences. Cambridge University Press, New York

Hausman DM, Woodward J (2004) Modularity and the causal Markov condition: a restatement. Br J Philos Sci 55(axh8):147–161

Hernan MA, Robins JM (2020) Causal Inference: What if. Chapman and Hall/CRC, Boca Raton

Illari P, Russo F, Williamson J (2011) Causality in the sciences. Oxford University Press, Oxford

Imbens GW, Rubin DB (2015) Causal inference for statistics, social, and biomedical sciences: an introduction. Cambridge University Press, Cambridge

Johansson LG (2019) An empiricist view on laws, quantities and physical necessity. Theoria 85(2):69–101

Menzies P, Price H (1993) Causation as a secondary quality. Br J Philos Sci 44(2):187–203

Pearl J (2000) Causality. Models, reasoning, inference. Cambridge University Press, Cambridge

Pearl J (2009) Causal Inference in Statistics: An Overview. Stat Surv 3:96–146

Phillips AW (1958) The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. Econ New Ser 25(100):283–299

Russell B (1913) On the notion of cause. Proc Aristotelian Soc 13:1–26

Samuelson P (1983) Foundations of economic analysis, enlarged edn. Harvard University Press, Cambridge, MA

Wittgenstein L, von Wright GH, Anscombe GEM (1969) Über Gewissheit: on certainty. Blackwell, Oxford

Woodward J (1997) Explanation, invariance, and intervention. Philos Sci 64:S26–S41

Woodward J (2016) Causation and manipulability. In: Zalta EN (ed) The Stanford Encyclopedia of Philosophy, winter 2016 edn. Metaphysics Research Lab, Stanford University

Chapter 7
Inferences from Statistics to Causation

Abstract Empirical results often consist of data organised as values of variables.The first question is whether an observed correlation is evidence enough for a correlation in the entire population. If the answer is yes, the next question is whether this correlation reflects a causal connection or not. That need not be the case, there might be a common cause. The main points of this chapter are:

- A correlation in an observed sample may not reflect any correlation in the entire population.
- There are three possible explanations for a correlation in a population: variable X is one cause of Y, or vice versa, or they have a common cause.
- The most reliable way to decide which is the case is to perform experiments.
- If experiments are impossible one can sometimes infer a causal relation from a natural experiment.

7.1 Inferences from Correlations and Regressions

Many sciences are full of correlations. The first question, when confronted with an *observed correlation*, is whether it is a random effect of the sampling or whether the observed correlation reflects a *true correlation in the entire population*. This distinction is very important to keep in mind when discussing possible causal links.

This question, whether there is a correlation in the entire population or not, cannot be answered with complete certainty. However, if the sampling is truly random, one can calculate a confidence interval for the correlation coefficient in the population, conditional on the observed one. The method is thoroughly described in textbooks on statistical inference. In Appendix C you will find an example of this calculation. *In what follows we take for granted that a correlation is observed in a sample, one has inferred that a correlation obtains in the entire population and this conclusion is correct.*

Why randomisation?

It was Fisher (1925) and Neyman (1923) who first stated that randomisation in sampling is necessary if one wants to make reliable inferences from a sample to the population. The reason is that one needs a probability distribution function when performing this inference, and which one should one choose?

One may conceive of the actual sample as one of numerous ones and if the sampling is random, the probability distribution for the mean values of these imagined samples is normally distributed with the same mean as that of the population. Furthermore, Fisher derived an equation relating the standard deviation of the sample s to the standard deviation in the population, σ. This means that if the sampling is random, one can use a normal distribution for calculating confidence intervals for mean values of different quantitative attributes. In particular, if we have a correlation between two variables in the sample one can calculate a confidence interval for the true correlation in the entire population.

Suppose we have done that and found a substantial correlation between two variables in the entire population. The next question is: what possible mechanisms can produce a correlation between two attributes of objects in an entire population?

In cases where the correlation is astonishing, given our background knowledge of nature and society, many people are inclined to conclude that the correlation must be a random effect. This is certainly possible, in particular if the correlation is observed in a small sample. But remember: in the following discussion about correlation and causation, the point of departure is that the inference from the correlation in the sample to the correlation in the entire population is correct.

Now, a correlation in an entire population consisting of an unlimited number of individuals *cannot be due to randomness*. This is a consequence of *the strong law of large numbers*, which is a theorem in statistics. It says, roughly, that if one randomly chooses a number of items from a population, the observed mean value of a stochastic variable in that sample will converge to the mean value of that variable in the population, when the number of items in the sample increases. So if we have a series of samples in each of which we observe a correlation between two variables, the observed correlation will approach the correlation in the entire population. So a correlation due to randomness may occur in a limited sample, in the procedure of selecting items for the sample, but not in the entire population. If there is a correlation in the entire population, we can reject the suggestion that it is a random effect. The question is then: how could there be a correlation in an entire, perhaps infinite, population?

The received view is, and we have no arguments to the contrary, that there are three possible ways for a correlation between two variables X and Y to occur in a population:

(i) X is one (direct or indirect) cause of Y;
(ii) Y is one (direct or indirect) cause of X;
(iii) X and Y have a common cause.

This is called *Reichenbach's principle*, after Hans Reichenbach (1891–1953) who first formulated it. It is easy to understand why these three types of mechanisms will produce a correlation. Could there be other mechanisms still? Not as far as we know.

How, then, do we decide which alternative is the case?

One can sometimes exclude alternative (i) or (ii), if one knows the timing of individual instances. If for example individual instances of X always occur in time before corresponding individual instances of Y, then one can exclude alternative (ii), and vice versa.

In some cases one can from a well established theory infer which alternative is the case. However, this is rarely the case in SES research; the field is still in its infancy and there are few if any well established theories in this field. Nevertheless one may sometimes guess that alternative (iii) ought to be the case, because any causal link between X and Y seems utterly implausible, given general scientific knowledge.

One example, although not from SES, is the strong correlation, 0.72, between prevalence of cousin marriage and percentage of wealth in cash, as measured across Italy's 107 provinces, see Fig. 7.1. There is no reason to believe that there is a direct causal link between these two features of Italian people's behaviour, so Henrich (2020) assumed that it must be a common cause, namely, people's degree of trust in foreigners. People with low trust in unknown and non-related persons are not inclined to invest their money in stock or put them in banks. Similarly, in communities with low trust in persons outside the extended family, marriage between unrelated persons are not popular and therefore is cousin marriage more prevalent. And conversely, people with high trust in other persons and institutions are more inclined to put their excess money in productive investments and are less sceptical to marriages with non-relatives. This is a plausible and testable hypothesis. Moreover, other studies have shown the same geographical variance in common trust over Italy's provinces. Roughly, the degree of trust is higher in northern than in southern provinces of Italy, whereas the proportion of cash and of cousin marriage in lower the more to the north an Italian province is situated.

Information about correlations is by itself seldom of any particular interest. Such information is a means to an end, the end of obtaining information about causal relations. To a great extent this interest is driven by our desire to act in the world: we try to prohibit unpleasant future events, if possible, or we try to increase the chance of future desirable events, if possible. In order to attain such goals, we need causal information: what should we do in order to bring about, or increase the chance,[1] of a certain effect? So we are looking for information about causal links and that is driven by our interests as agents in the world.

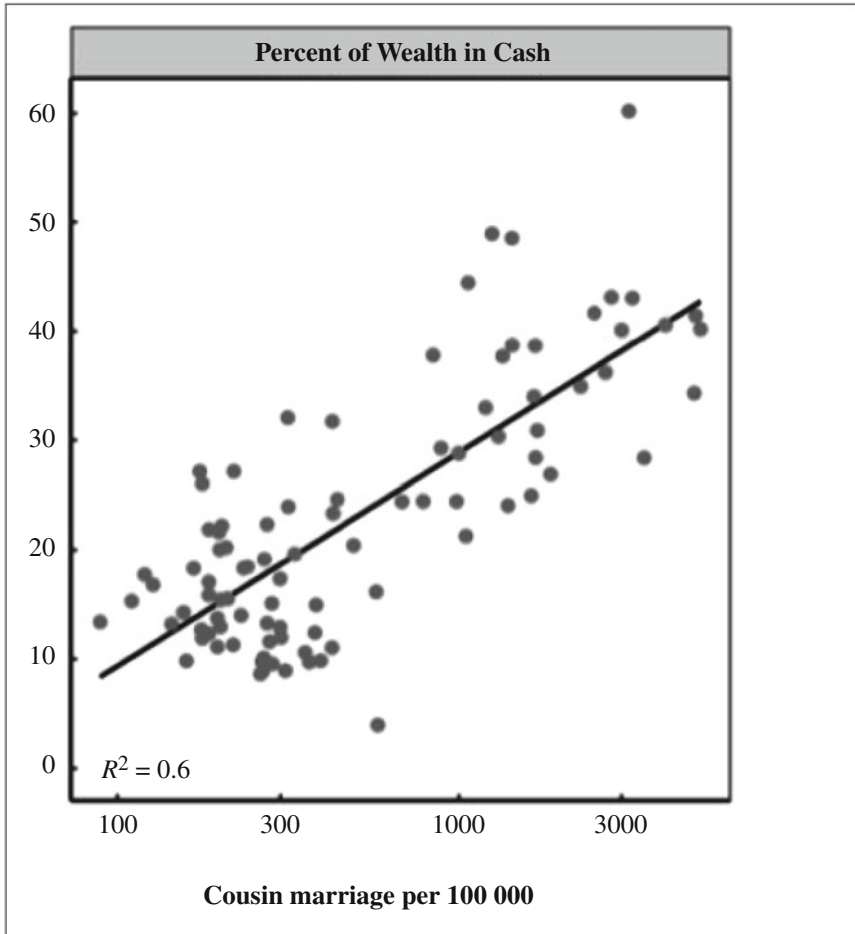[1] N.B. the causal terms 'bring about' and 'increase the chance'!

Fig. 7.1 The correlation between cousin marriage and percentage of wealth in cash in Italy's 107 provinces. Adapted from Blair Fix: Weird Consilience: A Review of Joseph Henrich's 'The WEIRDest People in the World', https://economicsfromthetopdown.com/2022/05/20/

It follows that the concept of cause is strongly related to the concepts of intervention and manipulability (cf. Sect. 6.7). We may cause a future event to occur, or at least increase its probability to occur, by doing something now. Or a present action may prohibit a future possible event, i.e., cause it *not* to happen.

It follows immediately that if we have a correlation between two variables X and Y and wonder whether X is a cause of Y (or vice versa, if Y-events precede corresponding X-events) we should manipulate X, i.e. make interventions, for example intentionally changing the values of the variable X and see if the values of Y changes concomitantly. This requires an experimental design.

Experimental tests is agreed to be the golden standard for testing hypotheses about causal relations. This nearly universal agreement about the optimal way of testing causal relations is no mere coincidence; it is a consequence of a core aspect of the meaning of expressions of the type '... is a cause of....'.

But what to do when experiments are impossible? This is very often the case in e.g., the social sciences such as economics and political science.

Correlations between economic variables are often observed and one may wonder which of all these correlations reflect causal links. It is often difficult, or impossible, to perform controlled experiments, both in macro and micro economics. There are at present two suggestions to obtain the needed information without performing carefully designed experiments: (i) observing *natural experiments* and (ii) controlling for *covariates*.

7.2   Natural Experiments

A natural experiment is not a consciously designed experiment, but a situation that in relevant aspects is similar to an experiment involving a test group and a control group. Here are two examples.

Example 7.1 Angrist and Pischke (2010, 13) discussed how to check the causal effect of class size on average test score in primary and secondary school. Does the size of the class, i.e. the number of pupils in a school class, have any causal effect on the average score among the pupils? Common sense has it that smaller classes leads to better scores, but in the data from US and many other countries there is no correlation between class size and score; sometimes it is even better scores in bigger classes. One cannot easily perform experiments, but the problem can be studied without conscious interventions, as is illustrated by the following case.

In Israel the class size is capped at 40, so if there are 41 students, these are divided into two classes each with circa 20 students. (Similarly, if there are 81 students, the group is divided into three classes, and so on.) One can then compare rather small classes with classes of around 40. Since the enrolment numbers to a particular school can be thought of as random, one has a situation sufficiently similar to one in which one performs a real experiment by randomly dividing schools in those with small classes and those with much bigger ones. In such circumstances one may assume that schools with different numbers of students per class are quite similar in other characteristics, hence if there is any difference in average scores, one may conclude that it is caused by differences in class size. And in fact there was a clear difference. Angrist & Pischke concluded: 'Regression discontinuity estimates using Israeli data show a marked increase in achievement when class size falls.' (op. cit. p. 14)

Example 7.2 The effect of informing US taxpayers that they had been fined for not carrying health insurance.

Here is a quote from Sarah Kiff in *New York Times*, Dec. 10, 2019, updated Dec. 13, 2019):

> Three years ago, 3.9 million Americans received a plain-looking envelope from the Internal Revenue Service. Inside was a letter stating that they had recently paid a fine for not carrying health insurance and suggesting possible ways to enrol in coverage.
>
> New research concludes that the bureaucratic mailing saved lives.
>
> Three Treasury Department economists have published a working paper finding that these notices increased health insurance sign-ups. Obtaining insurance, they say, reduced premature deaths by an amount that exceeded any of their expectations. Americans between 45 and 64 benefited the most: For every 1648 who received a letter, one fewer death occurred than among those who hadn't received a letter. In all, the researchers estimated that the letters may have wound up saving 700 lives.
>
> The experiment, an unintended result of a budget shortfall, is the first rigorous experiment to find that health coverage leads to fewer deaths, a claim that politicians and economists have fiercely debated in recent years as they assess the effects of the Affordable Care Act's coverage expansion. The results also provide belated vindication for the much-despised individual mandate that was part of Obamacare until December 2017, when Congress did away with the fine for people who don't carry health insurance. . . .

The budget shortfall mentioned was president Trump's decision to reduce the budget for IRS. It had the consequence that IRS stopped sending mails to those who had been fined for not carrying health insurance, so 600,000 uninsured individuals did not get any such letter. That enabled a comparison between sending and not sending such a letter, and that provided strong evidence for the conclusion that sending the letter caused a decrease in death rate.

7.3   Controlling for Covariates

Can one find out about causal relations without performing experiments and without access to information about natural experiments? Well, one can do one thing, namely, control for covariates.

The idea is that if the variable Z is a common cause of variables X and Y, we will observe that the correlation between X and Y disappears when we conditionalise on Z, which is feasible both for quantitative and category variables.

This is due to the fact that if X and Y are correlated (i.e., that the coefficient of correlation is far from zero), it holds that the joint probability P(XY) cannot be factorised. This means that either

$$P(XY) > P(X)P(Y) \text{ (positive correlation)} \tag{7.1}$$

or

$$P(XY) < P(X)P(Y) \text{ (negative correlation)} \tag{7.2}$$

while if

$$P(XY) = P(X)P(Y) \tag{7.3}$$

there is no correlation between X and Y.

If we have available values of a variable Z and conditionalise on it, we might find that

$$P(XY|Z) = P(X|Z)P(Y|Z) \tag{7.4}$$

i.e, that the joint probability for X and Y when conditionalised on Z is factorable. (In practice we will rarely find an exact equality. If the product $P(X|Z)P(Y|Z)$ is close to $P(XY|Z)$ the researcher may draw the conclusion that he has found the common cause.) If so, the variables $X|Z$ and $Y|Z$ are not correlated and this proves that Z was the common cause for X and Y.

But what if $P(XY|Z)$ is not factorable? This indicates that Z was not the common cause, or not the only common cause, there might be more than one common cause. Obviously, if there are several common causes and we only control for one of them we will not find that conditionalising on that one will result in factorability.

The difficulties in controlling for covariates is discussed in many papers. One useful contribution is (Witte and Didelez, 2019), where there are links to more literature on the subject. The abstract begins:

> When causal effects are to be estimated from observational data, we have to adjust for confounding. A central aim of covariate selection for causal inference is therefore to determine a set that is sufficient for confounding adjustment, but other aims such as efficiency or robustness can be important as well. In this paper, we review six general approaches to covariate selection that differ in the targeted type of adjustment set. We discuss and illustrate their advantages and disadvantages using causal diagrams.

The difficult question is of course how to discover all common causes when experiments are not possible. We can sometimes use well established theory, which gives us information of causal mechanisms. But this is no certain method, for how often can we be reasonably certain that our theory in relevant aspects is complete? In fact, if we were thus certain, we would not need any statistical analysis for determining whether a correlation indicates a causal link or not. Pearl (2000, 43) summarises our epistemological situation succinctly:

> In fact, the statistical and philosophical literature has adamantly warned analysts that, unless one knows in advance all causally relevant factors or unless one can carefully manipulate some variables, no genuine causal inferences are possible (Fisher, 1951; Skyrms, 1980; Cliff, 1983; Eells and Sober, 1983; Holland, 1986; Gärdenfors, 1988; Cartwright, 1989).

Suppose a researcher has discovered a correlation between two variables and has conditionalised on all factors that according to background scientific knowledge possibly could be linked to the two correlated variables. Let us further suppose that the correlation has survived this conditionalisation; does that prove that the there is a causal link between the correlated variables? No. Our scientific background

knowledge could be incomplete, there could be unknown common causes. There is no method for excluding this possibility. For if we had such a method, we could know whether our present best theory in a particular domain is complete or not. And we think that is in principle impossible.

Controlling for covariates can at most show that a correlation is not the result of a causal relation; but a positive proof of a causal relation is not possible. A thorough discussion about covariates and causal inferences can be found in (Waernbaum, 2008) and references therein.

7.4   Regression Analysis

Regression analysis is common and is often interpreted as giving information about the strength of causal relations. A linear regression of the form $Y = a + bX$, where a and b are constants is often interpreted as telling us that X is a cause of Y and b is a measure of the strength of the causal coupling. (Often the squared coefficient of correlation r^2 or the squared regression coefficient b^2 are used for giving a measure of the connection.) Thus Y is often called the response variable and X the explanatory variable. But as already pointed out, this cannot be inferred from the mere equation. It is obvious that the equation can be rewritten so that X is a function of Y. Hence, the distinction between explanatory variable (the cause) and response variable (the effect) must be based on some information not represented in the equation.

This is quite obvious from the fact that the correlation coefficient r_{xy} and regression coefficient b are related as

$$b = r_{xy} \frac{s_y}{s_x} \tag{7.5}$$

where s_x and s_y are the standard deviations in X and Y respectively. Since a correlation does not in itself tell us about any cause-effect relation, it is obvious that neither can information about a regression do that. *Regression* and *correlation* are statistical concepts; in order to make inferences about causal relations one needs additional information.

7.5   Heuristic: Hill's Criteria

Our theories about complex phenomena are mostly incomplete and experiments are often not possible. So one has to rely on uncertain indicators when trying to find out causes. Sir Austin Bradford Hill (1897–1991), who started epidemiology, proposed a set of nine criteria to provide epidemiological evidence for a causal relationship between a presumed cause and an observed effect, i.e., a disease, (Hill,

1965). In particular, he demonstrated the connection between cigarette smoking and lung cancer. (And when he was convinced that smoking was a cause of lung cancer, he stopped smoking!)

His list of criteria is as follows:

1. Strength (effect size): A small association does not mean that there is not a causal effect, though the larger the association, the more likely that it is causal.
2. Consistency (reproducibility): Consistent findings observed by different persons in different places with different samples strengthens the likelihood of an effect.
3. Specificity: Causation is likely if there is a very specific population at a specific site and disease with no other likely explanation. The more specific an association between a factor and an effect is, the bigger the probability of a causal relationship.
4. Temporality: The effect has to occur after the cause (and if there is an expected delay between the cause and expected effect, then the effect must occur after that delay).
5. Biological gradient (dose-response relationship): Greater exposure should generally lead to greater incidence of the effect. However, in some cases, the mere presence of the factor can trigger the effect. In other cases, an inverse proportion is observed: greater exposure leads to lower incidence.
6. Plausibility: A plausible mechanism between cause and effect is helpful.
7. Coherence: Coherence between epidemiological and laboratory findings increases the likelihood of an effect.
8. Experiment: "Occasionally it is possible to appeal to experimental evidence".
9. Analogy: The use of analogies or similarities between the observed association and any other associations.

As already pointed out, it is now generally agreed that careful double blinded experiments with control groups is the golden standard for inferring a causal relation between a manipulated variable and an observed variable which co-varies. In fact, this criterion virtually trumps all other factors mentioned by Hill. But in situations where experiments are impossible and where no natural experiment is available, one may use the other criteria for making informed guesses. Certainty cannot be expected, but an informed guess is better than nothing, see e.g., (Schünemann et al., 2010)

7.6   Summary

Scientists often report that they have observed an association between two variables. This word 'association' means the same as 'correlation', so they claim to have observed a statistical correlation. But in fact they have observed a correlation in a sample, for seldom, if ever, can one observe all items in an entire population. Saying that there is an association between two variables is in fact an inference from the sample to the entire population. This inference is always somewhat uncertain.

If in fact there is a correlation in the entire population and not only a correlation in the sample, one may ask how this correlation came about? What is the mechanism? There is general (albeit perhaps not completely universal) agreement that there are three possible types of causal mechanisms that can result in a correlation between the variables X and Y in an entire population:

  (i) X is one (direct or indirect) cause of Y;
 (ii) Y is one (direct or indirect) cause of X;
(iii) X and Y have a common cause.

Discussion Questions

1. Could there be any other explanations for a correlation than those mentioned in Reichenbach's principle?
2. Can you suggest some explanation, other than that given by Henrich, for the correlation between cousin marriage and percentage of wealth in cash?
3. Why are passive observations not sufficient for inferring causal connections in social science?

References

Angrist JD, Pischke J-S (2010) The credibility revolution in empirical economics: how better research design is taking the con out of econometrics. J Econ Perspect 24(2):3–30

Cartwright N (1989) Nature's capacities and their measurement. Clarendon Press, Oxford

Cliff N (1983) Some cautions concerning the application of causal modelling methods. Multivariate Behav Res 18:115–126

Eells E, Sober E (1983) Probabilistic causality and the question of transitivity. Philos Sci 50:35–57

Fisher RA (1925) Statistical methods for research workers. Oliver and Boyd, Edinburgh

Fisher RA (1951) The design of experiments, 6th edn. Oliver and Boyd, Edinburgh

Gärdenfors P (1988) Causation and the dynamics of belief. In: Harper W, Skyrms B (eds) Causation in decision, belief change and statistics, vol II. Kluwer, Dordrecht, pp 85–104

Henrich J (2020) The weirdest people in the world: how the west became psychologically peculiar and particularly prosperous. Farrar, Straus and Giroux, New York

Hill AB (1965) The environment and disease: association or causation? J R Soc Med 108(1):32–37

Holland PW (1986) Statistics and causal inference. J Am Stat Assoc 81:945–960

Neyman J (1923) On the application of probability theory to agricultural experiments. essay on principles. Roszniki Nauk Rolniczych 10:1–51

Pearl J (2000) Causality. Models, reasoning, inference. Cambridge University Press, Cambridge

Schünemann H, Hill S, Guyatt G et al (2010) The GRADE approach and Bradford Hill's criteria for causation. J Epidemiol Community Health 65(5):392–395

Skyrms B (1980) Causal necessity. Yale University Press, New Haven

Waernbaum I (2008) Covariate selection and propensity score specification in causal inference. Umeå Universitet, Umeå

Witte J, Didelez V (2019) Covariate selection strategies for causal inference: classification and comparison. Biom J 61:1270–1289

Chapter 8
Causal Explanations

Abstract There are several forms of explanation, one of which being causal explanations. Causal explanations are often descriptions of mechanisms, i.e., descriptions of how a state change in one object, labelled 'the cause', is transmitted through a number of intermediate objects to the final effect, i.e., a state change in another object. So the fundamental structure of mechanistic explanations is that of chained cause-effect relations.

The main points of this chapter are:

- Causal explanation is one kind of explanation beside several other kinds.
- A causal explanation often consists of describing the mechanism by which the cause produces its effect.
- Reasons for human actions are often viewed as causes of those actions, but that is controversial.
- Three types of causal explanations in terms of mechanisms are confounder mechanisms, feedbacks, and bifurcations.

8.1 Explanation-Seeking Questions

Scientific research aims at answering questions (Van Fraassen, 1980). As these answers accumulate, the knowledge grows, is corrected or completely revised. The starting point of research is usually a set of questions raised by practical concerns, or by earlier scientific research. Scientists' goal is to find convincing and correct answers to these questions. An important research skill is the ability to formulate these questions in a fertile manner. Part of this skill is the ability to break up the original set of questions into more concrete questions that can be answered by means of empirical research. Another part of the skill is research imagination, which allows the researcher to see which questions her methods and data can answer. Finally, the third part consists of the ability to design and conduct the research in a manner that convinces the audience that the answers presented by the researcher are better justified than competing accounts.

Most everyday research questions are descriptive. They ask about the facts regarding the world. Researchers seek to find out new facts about the world, but just as often they aim to check, correct or challenge what we believe to be facts. These questions usually start with words like Who? What? When? Where? Which? How many? How much? For example, the researchers studying Pleistocene extinctions ask questions like: Which mammal species went extinct during the late Pleistocene period? Where did these species live? What were their habitats? How large were their populations? When did they go extinct? Etc. These questions are often extremely tricky to answer conclusively.

Answering descriptive questions like these is the backbone of all scientific inquiry. But the scientific ambition is not limited to answering these questions; scientists also wish to answer explanatory questions. These questions often start with words like Why? or How? These questions take the answer to a descriptive question as their starting point and ask *why* that fact is the way it is. Many facts are puzzling to us, and we want to know why they happened or why they are one way rather than some other way. The fact to be explained is called the *explanandum*, and what explains it, the *explanans*.

Not all why-questions are explanation-seeking questions. Sometimes we ask 'Why?' when we want justification for a belief (Hempel, 1965) or for an action. For example, when someone claims that there was a mass extinction of large mammals during the late Pleistocene, it is reasonable to ask for some reasons to believe this claim. Because we do not yet believe the claim, we ask a justification-seeking why-question. However, once we take the claim to be a fact, we probably want to know what *caused* those extinctions. We want to understand why they happened. In this case, we would be asking an explanation-seeking why-question.

In the case of an explanation, the factuality of the explanandum is a presupposition of the explanation-demand, without which the question does not make sense. So, for example, if we ask (Barnosky et al., 2004), 'Why did most species of megafauna go extinct during the late Pleistocene (50,000–10,000 BC)?', we are presupposing that such a mega-extinction really happened. Sometimes the presuppositions of the question are not as obvious. For example, our question can be either read as assuming that there is a single cause for the mass extinction, or it can be read more loosely as allowing multiple independent causes. The first reading incorporates a quite strong assumption that easily can be false. It is possible that megafauna perished from different continents due to independent causes.

It is not always obvious that everybody shares the same presuppositions. A well-known anecdote about the famous 1930s bank robber Willie Sutton captures this. When a journalist asked Sutton why he robbed banks, Sutton responded, 'Because the money is there.' Clearly, he had a different contrast in mind than the journalist who was asking about his career choice.

One cannot expect that one answer to an explanation-seeking question could explain everything about the explanandum, e.g., the late Pleistocene extinctions. Typically we can explain only a certain aspect of a complicated event. A useful way to make the explanandum more precise is to articulate the intended contrast. The contrast describes an alternative state of affairs (the foil) that could have occurred

instead of the fact. For example, we could ask why did the extinctions happen during the late Pleistocene rather than some earlier or later period? Alternatively, we could ask why did the extinctions happen almost simultaneously rather than stretched over a longer period of time? We could also ask why the extinctions concentrate on megafauna rather than species of different sizes or all species? An explanans that provides an insightful answer to the last question might be quite uninformative about the other two questions, and vice versa. The contrast helps to pick up a causal difference-maker from the complicated causal history of Pleistocene ecology, and different contrasts can highlight very different difference-makers. Thus it makes sense to split the general explanation-seeking question into a series of more precise questions. The articulation of contrasts is a useful way to make explanation-seeking questions more precise (Van Fraassen, 1980; Garfinkel, 1981; Lipton, 1991).

Often our curiosity arises when we observe something unexpected, and we ask why things did not turn out as expected. We are quite curious when a person behaves in an unexpected manner, for example when he pours his coffee over his own head, but we are not usually asking for an explanation for his ordinary coffee drinking. The origins of our expectations might be in what we typically observe, theoretical predictions, or normative ideals. In everyday life, we usually explain surprising things, but in scientific research, also obvious things can become objects of curiosity (Hesslow, 1983). Why is the grass green rather than any other colour, or why there are two, rather than three biological sexes, are both meaningful scientific questions that are not raised outside science, except maybe by small children.

More generally, explanation-seeking questions are typical in theoretically oriented basic research. However, it would be a mistake to assume that such questions can be ignored by more practically oriented researchers. Reliable answers to such questions are usually descriptions of mechanisms and such descriptions are needed for the expansion of both theoretical and practical knowledge.

8.2 Explanations

The word 'explanation' can refer both to the activity of providing an explanation and to the product of that activity. While most discussions of explanation is focused on the latter, it is good to remember that explanation-seeking and explanation-giving are continuous social activities; we rarely provide complete explanations. Typical explanations, even in science, are limited by pragmatic contexts, they are more like sketches of explanations, leaving out relevant components that the reader can be assumed to be aware of in advance. Usually, in a given context we only highlight the salient features of the explanation and leave many background conditions unarticulated. This means that an explanation might have problematic presuppositions that we are not fully aware of. Furthermore, many of the explicit assumptions might be promissory: we believe that the facts that we have assumed are indeed the case, but we do not have sufficient evidence to support them. So,

if these presumptions turn out to be false, we have to reject or at least revise the explanation demand.

An answer purporting to be an explanation is expected to be true. An explanation that relies on false facts cannot be the proper explanation of an empirical observation. However, consisting of true statements is not enough, it also has to be relevant. First, it must answer the question by relieving the audience of the puzzlement the explanandum gave rise to (Lipton, 1991). Furthermore, this has to be done in a correct manner: it is not enough that the audience just thinks that they have understood or have a sense of understanding, as these metacognitive states are quite often unreliable. The audience might not get the explanation, for example, because it lacks sufficient background knowledge. It is also possible that an explanation is great in providing understanding but is unfortunately not true. Cases like this are called possible explanations (Hempel, 1965; Lipton, 1991). They are explanations that would have been satisfactory if their assumptions were true. Possible explanations are often an important element of the explanatory inquiry. For example, in the case of late Pleistocene extinctions, it is important to articulate a set of possible alternative explanations and then proceed to find evidence that discriminates between the alternatives (Barnosky et al., 2004; Stuart, 2014). Without the set of alternative explanations, we could easily mistakenly accept our first explanation as the correct one.

8.3 Different Kinds of Explanations

Answering a demand for an explanation is sometimes to provide a cause, or several causes, for the explanandum. But there are several other kinds of explanations that at least at first sight do not provide causes. We will here briefly discuss four such kinds before we delve into causal explanations: constitutive explanations, teleological explanations, functional explanations and intentional explanations.

8.3.1 Constitutive Explanations

In a constitutive explanation the capacities of a whole are explained by capacities of its components and their organisation (Ylikoski, 2013). The relation between the parts and the whole is not causal, hence this in not a case of causal explanation, which usually relates events to each other. However, it should be recognised that basically the same ideas about explanation apply to constitutive explanation that applies to causal explanation. Furthermore, constitutive explanation relates the causal capacities of the whole to the causal capacities of the parts, and, furthermore, changes in capacities are causal processes. So it would be highly misleading to say that constitution is completely unrelated to causation. Constitutive relations are an integral part of a causal picture of the world. The reason we have to recognise their

difference is that confusing part-whole relations with causal relations can lead to confused causal analysis.

8.3.2   Teleological Explanations

Another candidate for non-causal explanation is teleological explanation, which explains a process by an imagined goal, rather than by its causes. However, all forms of 'teleological' explanations in the sciences are actually subspecies of causal explanation (Elster, 1989). Biology is full of teleological explanations, but modern biology only accepts those that are supported by appropriate causal mechanisms. Natural selection is the prime example of such a mechanism. Consider, as an example, the human sclera, the white of the eye, which is a rare feature among great apes. According to the cooperative eye hypothesis, humans have white sclera because they facilitate telling the direction of gaze, which greatly facilitates non-verbal communication and coordination of action. The hypothesis explains the colour of the sclera by its beneficial consequences. However, for the hypothesis to be true, the claim has to be true about the past: the colour of sclera must be a heritable trait, and it must have given a relative fitness advantage to its carriers in earlier phases of human lineage because it facilitates cooperation. For example, if the colour is a by-product of some other trait, then the hypothesis is false. Thus when unpacked, the teleological claim is, in fact, a claim about a causal history.

8.3.3   Functional Explanations

Functional explanations are sometimes used in the social sciences. However, there is a quite broad consensus that they require an underlying causal mechanism. Finding such mechanisms are quite demanding, so proper functional explanations are quite rare in the social sciences.

In passing one may observe that term 'function' sometimes is meant to express a causal relation, sometimes only a mathematical or logical relation, the latter being common in natural and social sciences. One variable can be a function of another variable, but that can be the case without there being any causal link, as we discussed in the preceding chapter. This is not restricted to quantitative variables; if one boolean variable (i.e. having only two values, e.g. male-female) is correlated to another one (for example, do/do not enter higher education) one may correctly say that the second variable is a (probabilistic) function of the former one. But whether the first variable is a cause of the second one is a further question. If there is little or no evidence for there being a causal link between functionally related variables, one can hardly say that one explains anything just by pointing out that one is a function of the other.

8.3.4 Intentional Explanations

Finally, there are intentional explanations, which play an important role in the social sciences. Intentional explanations are not properly teleological explanations, because the explanatory work is done by a representation of a mental state (consisting of desires and beliefs) that precedes the outcome. It should be observed that the *mental state* of the actor precedes what is to be explained, while the *content* of that mental state is an imagined future event or state of affairs. Furthermore, the outcome causally depends on the agent having that mental state; if the mental state had been different, the action would have been different, which would have made a difference to the outcome. Intentional explanation has an extensive list of causal background conditions without which the connection between the mental state and the outcome will not hold. People cannot bring about, i.e., cause, things in the world just by having thoughts about them.

Intentional behaviour, i.e. actions, are usually explained by giving the agent's beliefs and desires; they are assumed to be the immediate causes of actions. From the point of view of everyday reasoning, the causal role of beliefs, desires and other mental states is quite obvious. We usually assume that beliefs and desires, i.e., reasons, can make a difference to the way we behave. Furthermore, in communication we attempt to influence each others' mental states and thus influence their behaviour.

The role of interpretive understanding in causal explanation of action highlights the importance of qualitative research. While much of it is descriptive, it describes what different people think, experience, and strive for, hence it lays the ground for causal explanations of their actions. To causally explain action, we have to get people's desires, feelings, and beliefs right. Similarly in institutional contexts we have to get right both the rules people follow and why they follow them.

8.4 Causal Explanation and Mechanisms

Apart from relevance and truth, an explanation requires the right kind of dependence between the explanans and the explanandum. Something is the case because it explains how facts are the way they are. For example, suppose that megafauna extinctions occurred because of human hunting. This is a claim about causal dependence: if there had not been extensive hunting of large mammals, the mass extinction would not have happened. Thus the explanatory claim is a claim about counterfactual dependence: if the cause had been different, the outcome would have been different too. Here the relevance criterion for the explanation is causal difference-making (Lipton, 1991; Woodward, 2003). While there is a huge number of things in the causal history of any event, the difference-making criterion helps us to pick the explanatorily relevant part of that causal history. If we have correctly

identified the right contrastive explanandum, we now have an informative answer to our explanation-seeking question.

While this kind of a simple causal statement might sometimes be enough to explain an event, we often want additional information. First, all causal claims hold only when certain background conditions hold (Mackie, 1974), see Sect. 5.6. If the background conditions do not hold, the cause would not be able to make the difference to be explained. So the first piece of additional information concerns the relevant background conditions. A better grasp of the background conditions helps us to see how the change we cite as the cause is embedded into a larger causal configuration. It might also help us to understand how fragile the causal connection is. It might well be that the cause can bring about the effect only in very rare circumstances. Understanding the relevant causal configuration helps to answer questions about the preconditions of the causal relation and possibly about alternative causes for the effect.

Another additional piece of information concerns causal mechanisms, that is, *how* the cause brought about the effect (Craver, 2007; Hedström and Ylikoski, 2010). This involves the idea that causation is a process, and describing that process increases explanatory understanding. One could say that information about the causal mechanism answers the how-question behind the causal why-question.

Knowledge of causal mechanisms is valuable for multiple reasons. First, evidence about mechanisms can help to justify the causal claim. A causal claim is more credible if there is a known mechanism by which the cause could bring about the effect and there is evidence that this particular mechanism has been present in the case at hand. Second, together with knowledge about the background conditions, understanding of the causal mechanism helps to understand how robust or fragile the causal relation is and what kinds of factors could prevent or modify the effect. Third, the mechanism helps to organise the causal explanation to a narrative that is easier to comprehend than individual claims about causal dependencies. Fourth, general knowledge in human and biological sciences is often formulated in the form of mechanism-schemes rather than general law-like generalisations. A mechanism-scheme outlines what kind of cause and causal configuration can bring about a certain type of effect. The outline has to be filled in for any particular explanatory use, but it provides useful guidance for the search for causes. It is often the case that there are alternative mechanisms that could bring about a similar effect. In cases like these, it is useful to have a toolbox of possible mechanisms that helps to find evidence that discriminates between alternative mechanistic scenarios.

Is there a general way to define what a mechanism is? In the literature there are many competing definitions. The entities and processes studied by different sciences are quite heterogeneous, so it is difficult to provide a definition that is both informative and covers all examples of mechanisms. One widely cited definition is the following:

> "A mechanism is a structure performing a function in virtue of its component parts, component operations and their organisation. The orchestrated functioning of the mechanism is responsible for one or more phenomena." (Bechtel and Abrahamsen, 2005, 423)

This definition might work in some areas of biology, but its application to, for example, SES research is difficult. While it is easy to recognise the importance of parts, operations, and organisation, the definition does not give much guidance for the construction of mechanism-based explanation. For example, it does not solve the problem of relevance: which entities, activities and their relations should be included in the explanation? A crucial aspect is that a mechanism-based explanation describes the causal process selectively. It seeks to capture the crucial elements of the process by abstracting away the irrelevant details. But how do we determine what is relevant and less relevant?

While a general definition is impossible, it is possible to say *something* general about mechanisms (Hedström and Ylikoski, 2010). First, a mechanism is always a mechanism for something, so the explanandum plays an important role in its identification. Second, *mechanism* is an irreducibly causal notion. It refers to the entities of a causal process that produces the effect of interest. A correlation between cause and effect is not enough for a mechanism, as it is based on the idea that there is a continuous process by which the causal influence is transmitted from the cause to the effect. Third, when a mechanism-based explanation opens the black box, it discloses this structure. In other words, it makes visible how the participating entities and their properties, activities, and relations produce the effect of interest. For this reason, the suggestion that a mechanism *just* is a chain of intervening variables misses an important point, each link in the mechanism must be a causal link. Conceptualising the mechanisms requires theoretical thinking. However, it also generates a series of additional hypotheses that can be tested, thus opening additional avenues for confirming causal claims.

While the idea of mechanism-based explanation is appealing, it is good to recognise some dangers associated with the idea. First, while explanation in terms of mechanisms comes naturally to us, we quite often end up with mechanistic storytelling. This means that we are satisfied with the first sketchy mechanism-story that we could come up with and do not bother to consider alternative mechanisms or to check whether our story agrees with empirical evidence. Second, quite often people just name a mechanism rather than describe how it is supposed to work. While this kind of intellectual laziness is understandable, it is not supported by the core idea of mechanism-based explanation. The goal of mechanism-based theorising is not to create illusory understanding, but to fight it.

8.5 Some Special Mechanisms

Causal mechanisms can consist of several different types of structures. Here we will briefly discuss confounder mechanisms, feedback mechanisms and bifurcations, which are of particular interest in SES.
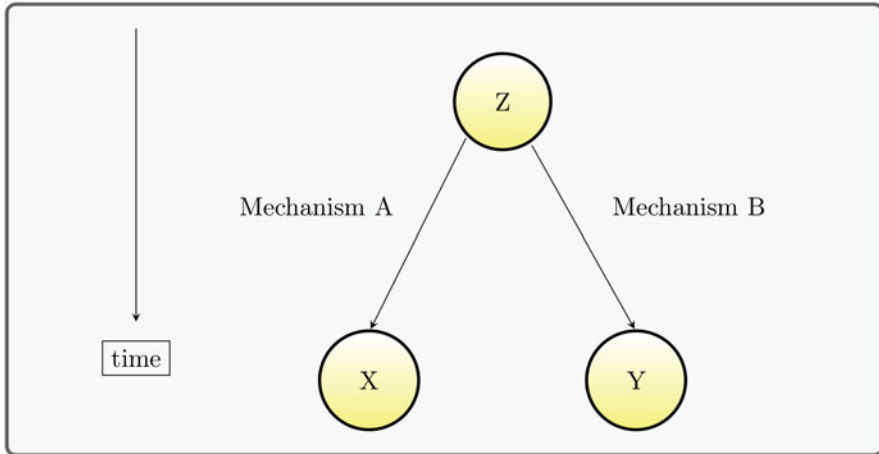
Fig. 8.1 The variables X and Y are correlated because both depend on a common cause Z. N.b: there is no arrow between X and Y, since there is no causal mechanism going from X to Y!

8.5.1   Confounder Mechanisms

Confounders are common in empirical sciences. A confounder is a non-observed variable which is a common cause of two observed and correlated variables. A strong correlation between two observed variables can be due to three possible causal connections, according to Reichenbach's principle (see Sect. 7.1); variable X is one the causes of variable Y, or vice versa, or there is an unobserved common cause Z. This is often called a confounder.[1]

The two mechanisms where a common cause, the confounder, produces a correlation between the two observed variables X and Y can be visualised by the following directed graph (Fig. 8.1):

The structural equation system for this situation is:

$$\begin{cases} X =: k_1 Z \\ Y =: k_2 Z \end{cases}$$

where k_1 and k_2 are the coefficients of correlation, if there are no other common causes of X and Y. Since correlation is a transitive relation the correlation between X and Y is the product $k_1 k_2$. We may thus infer that in order to observe even a weak correlation of e.g., r = 0.3 between X and Y the coefficients k_1 and k_2 must be rather substantial.

---

[1] It is possible, and perhaps rather common, that a correlation between two variables X and Y are due both to there being a causal link between them and them having a common cause.

This in turn means that if there are unknown causes other than Z independently affecting X and Y, the coupling coefficients k_1 and k_2 will be weak, hence the correlation between X and Y will in such cases be very weak and often not discernible.

The time arrow must be interpreted with caution. It tells us that *individual events* of the type 'variable Z has the value z_i' occur earlier than the individual events of the types 'variable X has the value x_i' and 'variable Y has the value y_i'. But the variables themselves, which are mappings from events to numbers, cannot be attributed times, since they are abstract entities.

8.5.2 Feedback Mechanisms

Feedback mechanisms are common in complex systems, see e.g. the following quote:

> The essential element in any SES with persistent structure (e.g., an ecological community and its human dependents) is feedback (Csete and Doyle, 2002; Carlson and Doyle, 2002). In SESs, these feedbacks take the form of information-action loops wherein human individuals or groups extract information about the state of a system (e.g., an ecosystem), decide how to act on the system (e.g., which species to protect and which to harvest), and undertake the action, generating a response from the ecosystem (e.g., changing population size or distribution), that over time triggers system change and restarts the cycle (loop) (Anderies et al., 2007, 2019). (Anderies et al., 2022a, 3)

As discussed in Sect. 5.4, a feedback mechanism does not contradict the fundamental idea that an individual effect cannot precede its cause. When we talk about feedback mechanisms we always talk about relations between variables. Variables are abstract entities, sets of values (in the case of quantitative variables), or attributes of concrete events, objects or states of affairs (in the case of category variables). To repeat, abstract things such as sets or attributes do not exist in space and time. Hence in Fig. 8.2 there can be no time line.

But we have a natural tendency to interpret figures of this type as if there are time relations between the items, and taking arrows as indicating processes in time. *This is not correct for feedback diagrams*! We have discussed use of diagrams more extensively in (Banitz et al., 2022b).

The structural equation system (cf. Sect. 6.5.2) for this case is

$$\begin{cases} X =: k_1 Z \\ Y =: k_2 X \\ Z =: k_3 Y \end{cases}$$

A variation of dz in Z will result in a variation $k_1 dz$ in X, which will produce a variation $k_1 k_2 dz$ in Y, etc., hence the 'strength' of the feedback mechanism is the product $k_1 k_2 k_3$. It is obvious how to generalise to any number of intermediate
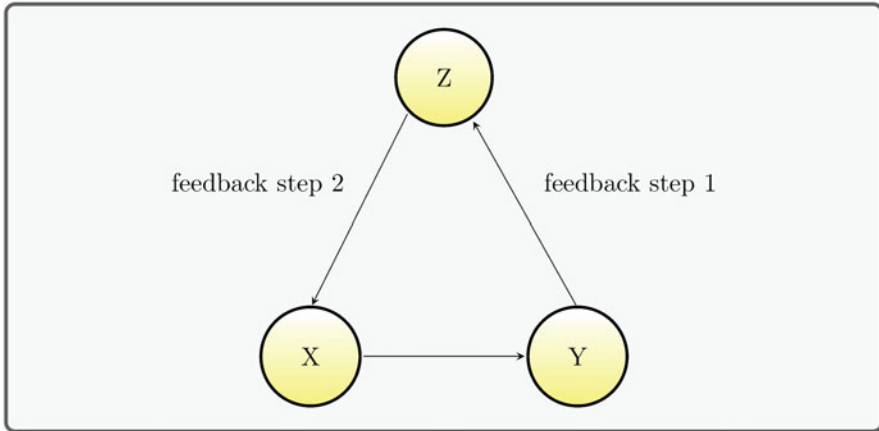
Fig. 8.2 The feedback from Y to X goes via the variable Z

variables making up the feedback. Furthermore, if we have observed a system with an effective feedback, each of the three coupling coefficients must be high. It follows that if one is able to manipulate one of the couplings by an intervention one might break down the feedback.

8.5.3   Bifurcation Mechanisms

Functional relations (see Sect. 5.2.2) state not only that there are relations between variables, but also provide a detailed account of properties of these relations. Take for example a system of ordinary differential equations that describes a freshwater lake ecosystem exposed to nutrient runoff:

$$\frac{db}{dt} = r_b \frac{n}{n + h_0} b - c_b b^2 - k_1 \frac{b^2 p}{b^2 + h_1^2},$$

$$\frac{dp}{dt} = k_2 \frac{b^2 p}{b^2 + h_1^2} \frac{v}{v + h_2} - c_p p^2 - m_p p. \tag{8.1}$$

Letters b, p and v denote bream, pike and vegetation respectively and represent the key species in the lake ecosystem. Parameters in differential equations, denoted by k_1, h_1 and k_2 define the strength of interactions between bream and pike, parameters r_b, c_b, c_p, m_p define ecological processes of each species and parameter n define bream response to the amount of nutrients in the lake water.

Figure 8.3 shows changes in the lake ecosystem dynamics due to changes in the amount of nutrients (i.e. values of parameter n). For small values of parameter n,
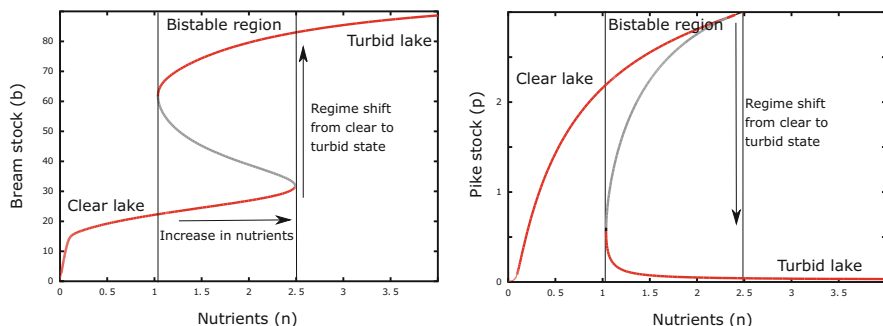
Fig. 8.3 Bifurcation diagrams for bream and pike in relation to the nutrient level. Red lines represent stable regimes, grey lines are unstable states. Arrows indicate nutrient increase and consequent regime shift from clear to turbid state

the lake water is clear and bream levels stay low, but more intensive nutrient load (represented by higher values of n) can lead to eutrophication of the lake, increased bream levels, decreased pike levels and changed structure and functioning of the ecosystem. This creates a bistable region, where depending on the initial conditions, the lake can evolve toward clear or turbid state. Further increase in nutrients leads to turbid lake state. The shift from clear to turbid lake state due to increase in nutrient load (and parameter n) is an example of a *regime shift*, a phenomenon that can be explained by *bifurcation mechanisms*.

Bifurcation mechanism means that qualitative properties of system dynamics change due to changes in strength of individual interactions or drivers. We have discussed this at some length in (Radosavljevic et al., 2023).

8.6   Summary

Explanations in general, and scientific explanations in particular, are highly context-dependent because a number of background assumptions are usually made without being explicitly stated. Scientific explanations are in most cases explanation sketches, not complete explanations.

There are several types of explanations, one of which being causal explanations. Teleological and functional explanations are, on closer inspection, causal explanations.

Causal explanations are often given by describing a mechanism which tells us how the cause produces its effect.

Discussion Questions

1. How do you respond to an explanation request that has a tacit and false background assumption?

2. What is the difference between functional description and functional explanation?
3. Is there a way of describing a causal mechanism that is not built out of chains of variables coupled via structural equations?
4. Is it possible to discern any ultimate constituents of a chain of elements constituting a mechanism, constituents that need no further analysis? If so, why are these the 'bottom'?
5. If a person believes that she will achieve a goal G by performing a certain action A, and if that person desires G, she will perform action A. Do you think that an explanation of this person's action A consists of a description of this particular belief and desire? If so, is it a causal explanation?

References

Anderies JM, Rodriguez AA, Janssen MA, Cifdaloz O (2007) Panaceas, uncertainty, and the robust control framework in sustainability science. Proc Natl Acad Sci 104(39):15194–15199

Anderies JM, Mathias J-D, Janssen MA (2019) Knowledge infrastructure and safe operating spaces in social-ecological systems. Proc Natl Acad Sci 116(12):5277–5284

Anderies JM, Cumming GS, Clements HS, Lade SJ, Seppelt R, Chawla S, Müller B (2022a) A framework for conceptualizing and modeling social-ecological systems for conservation research. Biol Conserv 275(Nov):109769

Banitz T, Hertz T, Johansson L-G, Lindkvist E, Martínez-Peña R, Radosavljevic S, Schlüter M, Wennberg K, Ylikoski P, Grimm V (2022b) Visualization of causation in social-ecological systems. Ecol Soc 27(1)

Barnosky AD, Koch PL, Feranec RS, Wing SL, Shabel AB (2004) Assessing the causes of late pleistocene extinctions on the continents. Science 306(5693):70–75

Bechtel W, Abrahamsen A (2005) Explanation: a mechanist alternative. Stud Hist Phil Sci Part C Stud Hist Phil Biol Biomed Sci 36(2):421–441

Carlson J, Doyle JC (2002) Complexity and robustness. Proc Natl Acad Sci 99(Suppl. 1):2538–2545

Craver C (2007) Explaining the brain: mechanisms and the mosaic unity of neuroscience. Oxford University Press, Oxford

Csete ME, Doyle JC (2002) Reverse engineering of biological complexity. Science 295:1664–1669

Elster J (1989) Nuts and bolts for the social sciences. Cambridge University Press, Cambridge

Garfinkel A (1981) Forms of explanation : rethinking the questions in social theory. Yale University Press, New Haven

Hedström P, Ylikoski P (2010) Causal mechanisms in the social sciences. Ann Rev Sociol 36(1):49–67

Hempel C (1965) Aspects of scientific explanation. The Free Press, New York

Hesslow G (1983) Explaining differences and weighting causes. Theoria 49(2):87–111

Lipton P (1991) Inference to the best explanation. Routledge, London

Mackie J (1974) The cement of the universe: a study of causation. Clarendon Press, Oxford

Radosavljevic S, Banitz T, Grimm V, Johansson L-G, Lindkvist E, Schlüter M, Ylikoski P (2023) Dynamical systems modeling for structural understanding of social-ecological systems: a primer. Ecol Complex 56(Dec)

Stuart AJ (2014) Late quaternary megafaunal extinctions on the continents: a short review. Geol J 50(3):338–363

Van Fraassen BC (1980) The scientific image. Oxford University Press, Oxford

Woodward J (2003) Making things happen: a theory of causal explanation. Oxford University
    Press, Oxford

Ylikoski P (2013) Causal and constitutive explanation compared. Erkenntnis 78:277–297

Part III
Causation in Complex SES

Chapter 9
Causation in Social-Ecological Systems Research

Abstract The book has so far introduced fundamental ideas about causation, i.e., the relation between cause and effect, from philosophy, particularly those ideas that underlie studies of causation based on quantitative data and statistical methods of causal inference (Chaps. 1–7). Knowledge of these concepts, ideas and associated methods is essential as they are often used in sustainability science studies rooted in the natural sciences, economics and other quantitative social sciences. The book has also introduced the notions of causal explanation and causal mechanisms, which are used more broadly in both quantitative and qualitative studies to explain how a cause brings about an effect (Chap. 8). In this last chapter we want to reflect on causal reasoning from a broader angle, to illustrate the diversity of ways in which sustainability researchers reason about causation, and to highlight the many instances within a research process in which researchers engage in causal reasoning.

9.1 Introduction

Causal reasoning, as we define it here, refers to the cognitive activities we engage in when figuring out the effects of specified causes, and how these effects are brought about, but also when identifying causes that may produce specified effects. These activities usually involve making sense of the broader causal setting, i.e., the context, in which the causes and effects of interest operate. It also involves choosing a research design and methods, collecting or generating data, interpreting results and, finally, justifying causal claims.

Causal reasoning thus does not only occur when we interpret data to test whether there is a causal relationship; it takes place throughout the entire research process and it can be very diverse. The potential outcomes framework introduced in Chap. 4 (Rubin, 2004, 2005) which is currently receiving lots of attention (e.g., Kimmel et al., 2021), is only one way to reason about causal relations. It is rooted in a particular idea of causation that is associated with the experimental method, and puts much emphasis on the design of a study, because the quality of the design determines the ability to make a causal claim. There are also other ideas about causation, which have been formalised to various degrees. The potential outcomes framework is

formalised mathematically, other ideas about causation less so. Different approaches also put different emphasis on the concept of a causal relation, which methods are suitable, and what is considered appropriate evidence for a causal claim (Illari and Russo, 2014).

In addition, different approaches may differ in how researchers build their causal models, and whether they search for causes in a single case or for general causes across a population of cases, cf. Sect. 2.3.

This chapter aims to introduce this broader picture and invites the reader to explore the variety of ways sustainability researchers reason about causation. An understanding of this diversity and ways to navigate it is important for inter- and transdisciplinary collaboration and for assessing causal claims and their consequences for action. Both collaboration and reflexivity are critical for enhancing understanding of social-ecological systems and for finding appropriate solutions for sustainability problems. To this end, we discuss how causal reasoning proceeds in a study, illustrate the diversity of causal reasoning with some examples and conclude with pointing to some tools and further readings that help to clarify causal reasoning.

Making sense of and analysing causation in complex social-ecological systems (SES) is an emerging research frontier. This chapter provides some initial ideas, but a more thorough treatment is beyond its scope. For a deeper exposure to particular aspects of this broad frontier, we refer the reader to the literature (cited throughout the text and in the suggested readings below).

9.2   Causal Reasoning About Social-Ecological Systems

Researchers study social-ecological systems with the aim to enhance understanding of pressing environmental problems and potential solutions to address them (Kates, 2011). We want to understand what causes a problem, such as biodiversity loss, or the deterioration of a freshwater lake and what can be done about it. The field is inter- and transdisciplinary. It involves several disciplines from the social and natural sciences and the humanities and it co-produces knowledge with practitioners and stakeholders in participatory and change-making processes (Lang et al., 2012; Norström et al., 2020; Chambers et al., 2021). Both understanding and action are thus key goals in SES research and are intricately linked. Pursuing both goals requires causal reasoning. This causal reasoning can vary significantly among diverse actors that bring their different backgrounds, experiences and values to the study of SES and their search for solutions. How to deal with the plurality of causal understandings and the co-production of causal knowledge in inter- and transdisciplinary processes is an important challenge and research frontier (Schlüter et al., 2023b; Caniglia and Schlüter, 2023).

Researchers with different backgrounds bring different world-views and epistemologies to the study of causation in complex SES. Causal questions such as whether a cause produces the intended effect, which causal processes have generated an outcome of interest, or, more generally, questions about how SES work, will be

answered differently across diverse disciplines. This is so because they build on different beliefs about what should be considered a cause-effect relationship, they differ in what kind of relations are considered interesting and important or they differ fundamentally in their views of the world.

Different research traditions also have different normative standards about what counts as acceptable evidence for causal claims, how this evidence should be collected and what can be generalised from particular studies. These norms are associated with certain epistemologies and preferences for approaches and methods, such as viewing experiments as the golden standard for causal inference, versus viewing in-depth historical studies that trace causal pathways as the best way to understand causation.

Bridging these different views and approaches into dialogue in ways that respects their differences and bridges them where possible is important because no single approach to causation can deal with all problems or aspects of SES. The complex, multi-scalar and social-ecologically intertwined nature of SES pushes the limits of the reasoning and methods used. In addition, the field requires approaches that move beyond a conception of linear causality towards conceptions that acknowledge the complex nature of SES (Preiser et al., 2021; Geels, 2022).

One way sustainability researchers have dealt with the complexity of sustainability problems and the challenges of interdisciplinary collaboration is through the construction and use of frameworks, which has led to their proliferation (Biggs et al., 2022, ch.3). Frameworks are collections of concepts that are considered to be most relevant for studying a particular phenomenon. Frameworks often also include ideas about causal relations, e.g., how a change in one element affects another element. A prominent example of a framework to study collective action is the SES framework proposed by Ostrom (2007). That framework has been developed to provide a comprehensive set of variables that have empirically proven to be relevant for explaining cases of successful collective action for managing a common pool resource. Ostrom's framework refrains from specifying causal relationships, this is the function of theories. However, it does assume relationships between variables at the highest level, e.g. that the resource system, resource units, governance system and users have a direct causal influence on interactions and outcomes. An overview of the most common frameworks in SES research can be found in Biggs et al. (2021).

9.3   Causal Reasoning in a Study

When we conduct a study, the causal reasoning we use is shaped by processes that take place outside the framework of the study, because the research context and the scientific and practical backgrounds of those involved (Fig. 9.1; Box 1, 2) set the stage for the study. The causal inferences made in the study are influenced by the participants' causal understanding of the social-ecological system in which the phenomenon is embedded (Fig. 9.1; Box 2). This, in turn, is influenced by disciplinary backgrounds, experiences, literature, selected theories, frameworks,
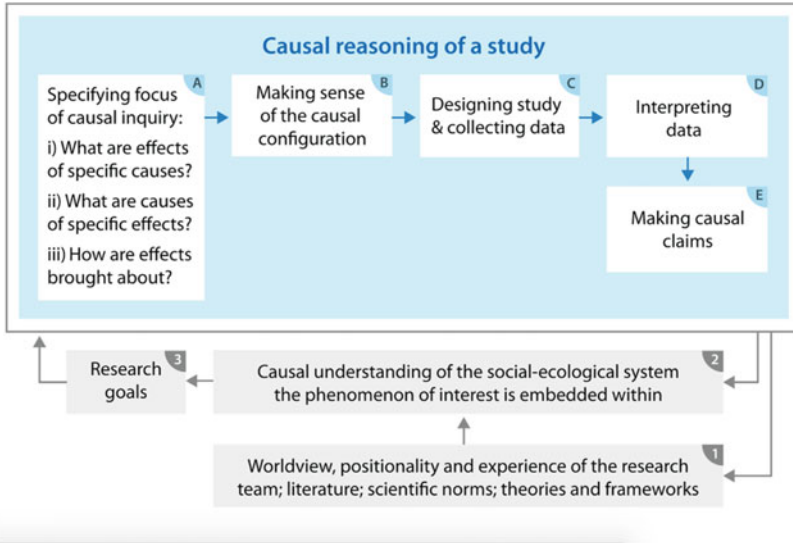
Fig. 9.1 Different instances of causal reasoning during a research process. Causal reasoning occurs during different stages of a research process. It is influenced by the worldview, positionality, and experiences of those involved in the study. Literature, scientific norms, theories and frameworks also influence causal reasoning and the prior causal understanding of the phenomenon of interest and the social-ecological system in which it is embedded (Box 1, 2). The research goals (Box 3) influence the causal questions a study asks, e.g. whether the aim of the study is to measure the effects of specific causes or identify the causes of specific effects (Box A). A study can address several of these questions. Once the goal of the causal inquiry has been set, the next step involves making sense of the causal configuration (Box B). This informs the design of the study and data collection or generation (Box C). The final steps are taken when researchers interpret the results (Box D) and justify their causal claims (Box E). The new understanding of the system and the causal configuration gained may feed back into the broader context of the study (Box 1, 2)

and scientific norms of what is acceptable and desirable in scientific practice in a given community. Furthermore, participants' world-views, positionalities (i.e., gender, cultural background, class, country of origin) and everyday experience (Fig. 9.1; Box 1) informs and motivates research goals.

The goal of a study, e.g. whether the aim is to predict (what may happen in the future?), intervene (what is the best way to bring about a desired effect?), explain (why and how did something happen?), or attribute responsibility (what cause was decisive in bringing about an effect?) shapes the subsequent causal reasoning.

The research goals also influence the focus of the causal inquiry, i.e. what kind of causal questions will be prioritised, e.g. whether a study focuses on the effects of specific causes, how effects are brought about, or the causes that bring about specific effects (Fig. 9.1; Box A. i–iii). The goals and questions, together with the background understanding and position of the researcher (Fig. 9.1; Box 1–2) influence how researchers make sense of the causal configuration of the phenomenon of interest (Fig. 9.1; Box B).

9.4   Causal Configuration

Social-ecological phenomena are complex, they are composed of a variety of elements and interactions that are organised in a specific way in both time and space. This is the causal configuration of the phenomenon of interest.

Researchers normally begin a study with a mental model of the causal configuration of interest (Fig. 9.1; Box B). This model is informed by the researcher's prior understanding of the system at hand.(Fig. 9.1; Box 2). Then new knowledge is generated, resulting in an updated model of the causal configuration.

For example, if we are interested in the governance of an eel fishery, prior knowledge of local institutions, eel biology, fishing styles, and changes in landings informs our mental model of the causal configuration. We then learn new details about the causal configuration through the study, such as the diversity of fishers' livelihoods and adaptation strategies to financial and climatic shocks, competition, and incentives. This new knowledge results in a more elaborated model of the causal configuration.

The representation of the causal configuration made by the research or co-production team, their methodological and theoretical background and data accessibility, inform the selection of methods, possible intervention, and data collection (Box C). The design of the study and the methods used strongly influence the causal interpretation of the results.

9.5   Interpreting Results

After data are obtained and processed (Fig. 9.1; Box C), causal reasoning focuses on interpreting data as evidence (Fig. 9.1; Box D). This depends on background information about the causal configuration (Fig. 9.1; Box 2) and is a critical step to figure out whether the data is evidence of a causal relation or not. Scientists might give different reasons, provide different interpretations, or favour one instead of another. However, it is possible to identify some commonly used schemes of reasoning supporting causal conclusions, such as (i) that the cause precedes the effect (cf. Sect. 3.4), (ii) that a correlation is an indicator of a causal relation (cf. Sects. 7.1–7.3), (iii) that the cause and the effect are linked through a mechanism (cf. Sect. 8.4), (iv) that manipulating the cause will change the effect in otherwise invariant conditions (cf. Sects. 6.1, 7.1 and 7.2), or (v) that the most likely causal explanation is the one that best makes sense of all the available evidence (cf. Sect. 8.4).

9.6   Making and Justifying Causal Claims

The final stage of a causal study is to write a report where the conclusions are drawn and the arguments for these conclusions are given (Fig. 9.1; Box E). The reasoning at this stage builds upon the prior stages of the study, but often reshapes and refines it. The crucial point of this stage is to provide justificatory support for claims.

The strength assigned to a causal claim should match the support provided by the evidence. For instance, when claiming a causal relation between two (quantitative or qualitative), variables, it is not enough to refer to an observed correlation between them. At most one may claim that there *might* be a causal relation. Causal claims might vary regarding their strength, specificity, and scope. Compare these four claims:

- Baseline claim. High trust among male small-scale fishers in Kino bay, Mexico, seems to lead to an increase in their income.
- Stronger claim. High trust among male small-scale fishers in Kino bay, Mexico leads to an increase in their income.
- More specific claim. High trust among male small-scale fishers in Kino bay, Mexico, leads to an increase in their income of 20%.
- Wider scope claim. High trust among small-scale fishers around the world seems to contribute to their income.

   Each of these claims require different kinds of evidence.

   When communicating research or interpreting other people's research one needs to be aware that, claims are differently justified and differently interpreted. In inter- and transdisciplinary spaces there can be tensions between the evidence provided for causal claims and scientific standards for justification of causal claims.

9.7   Diversity of Causal Reasoning

Depending on the goals of a study, the problem to be investigated and the chosen approach, each case of causal reasoning outlined in Fig. 9.1 will be unique. To illustrate this diversity, we explore causal reasoning in five exemplary studies from SES research. Example 1 is a case of statistical causal inference that examines whether a community monitoring program can reduce groundwater extraction from aquifers, improve water quality, and increase user satisfaction in Costa Rica (Carpio et al., 2021). Example 2 aims to explain the synchronicity of recent global crises, such as the 2008 food-energy crisis and the financial-energy crisis (Homer-Dixon et al., 2015). Example 3 examines the case of the Baltic cod collapse (Lade et al., 2015). Example 4 studies the mechanisms that may explain the emergence of self-governance arrangements in fisheries in Mexico (Lindkvist et al., 2017) and example 5 examines how a practice-based approach to sustainability interventions can support workable solutions in ever-changing contexts (West et al., 2019).

9.7.1 Study 1: Quantifying the Effect of Community-Based Monitoring on Groundwater Management: A Statistical Causal Inference Approach

The goal of Carpio et al. (2021) was to investigate whether there is a causal relation between an externally driven community monitoring program and improved groundwater management in rural Costa Rica, and if so, to quantify the effect (Fig. 9.1; Box 3). It thus asks the causal questions 'what are the effects of a specified cause, i.e., the community-based monitoring', and 'what are the magnitudes of these effects and how are they brought about?' (Fig. 9.1; Box A).

The authors develop their understanding of the causal configuration that underlies the effect of community-based monitoring on groundwater management (Fig. 9.1; Box B) using literature from three empirical and theoretical fields: common pool resources, community-based environmental monitoring and citizen monitoring of public services (Fig. 9.1; Box 1, 2). This knowledge was used to specify a hypothesised mechanism through which monitoring (i.e., interventions) influences the quality of water management. The causal configuration informed the development of three hypotheses about the effects of monitoring.

These hypotheses were tested using a randomised experimental design where the causal variable, i.e. community-based monitoring, was externally manipulated through applying an intervention to some communities but not others (Fig. 9.1; Box C). This approach assumes that through manipulating the community-based monitoring (the assumed cause) we can obtain knowledge about its connection to the assumed effect, and that randomisation eliminates the influence of contextual variables and makes the communities comparable. The monitoring intervention was applied to communities that were randomly selected, but not to those in the control group and data on the primary outcomes and intermediate variables were collected (Fig. 9.1; Box C). The data was then interpreted using counterfactual reasoning, i.e., the changes in outcome variables between treated and control units were compared (Fig. 9.1; Box D). Final and intermediate outcomes that are part of the mechanisms were also measured and analysed statistically (cf. Sect. 6.7: Causation, Manipulation and Intervention).

The experimental results provided some evidence for the causal claim that community monitoring improves groundwater management (Fig. 9.1; Box E) because the impacts of the intervention point in the right direction (communities with monitors pumped less, had better water quality and higher customer satisfaction), but impacts after 1 year of the program were modest. The authors also found evidence consistent with their theory of change, but the effects of the program on the intermediate outcome variables were small and imprecisely estimated. The conclusions were justified by experiments and the specification, and partial verification, of a plausible mechanism (cf. Sect. 8.4: Causal Explanations and Mechanisms).

However, no alternative mechanism was discussed. In their discussion, the authors reflected on factors that could make the intervention more successful, e.g. why a particular causal pathway was not very strong and how it could be strengthened, and on the implications of the results for action. There is no discussion on how this study may have changed the causal understanding of the system or phenomenon of interest.

9.7.2   Study 2: Synchronous Failure: The Emerging Causal Architecture of Global Crisis

The researchers who conducted this study (Homer-Dixon et al., 2015) were interested in explaining the synchronicity of recently emerging world crises (Fig. 9.1; Box 3). Since we, the co-authors of this book, did not conduct the study, we can only speculate about how the authors' personal trajectory shaped their understanding of the system prior to conducting the study (Fig. 9.1; Box 1 and 2).

This study has two parts, in the first part authors constructed a plausible causal model of world-scale crisis synchronicity, and in the second they validated the causal model with empirical evidence from case studies.

The focus of this causal inquiry was on how synchronicity of global crises emerge (Fig. 9.1; Box A.ii). To build the model of the causal configurations responsible for this outcome, authors looked at processes that have shaped human-nature interaction during the last decades. They argued that, as the scale of human activity, resource use, and world connectivity has increased, the flows of information, matter and energy between subsystems have become more intense, as well as their proneness to crises. Then, the authors represented these features in three stylised and interconnected models inspired by complexity theories.

As an example, the long fuse big bang captures the non-linear behaviour and configurational change of world subsystems—like the energy, food or economic subsystems—when their coping capacity is exceeded. Simultaneous stresses on subsystems erodes their capacity to endure stress, which eventually leads to a big bang and ramifying cascades, which captures the way in which crisis propagation happens across interconnected subsystems.

Overall, their causal model proposes that the synchronicity of world crises is a consequence of three factors: (i) the simultaneity of stresses across world subsystems, (ii) their homogeneous proneness to crisis, and (iii) the tight connectivity that allows for crisis propagation (Fig. 9.1; Box B). This hypothetical model is an example of reasoning in terms of INUS conditions (cf. Sect. 5.6).

In the second part of the argument, the authors looked at two case studies of simultaneous global crises, the 2008–2009 food-energy crisis and the financial-energy crisis in the same years (Fig. 9.1; Box C). The authors interpretation of these case studies consisted in mapping them out onto their proposed model (Fig. 9.1; Box D). The model advanced by the authors at the beginning of the argument was

not changed, but it was updated in regards to its empirical support; the authors claimed that 'recent global crises reveal an emerging pattern or architecture of causation that will increasingly characterise the birth and progress of crises in the future'. This claim gets justificatory support from the plausibility of the model of synchronous crises, it's consistency with theories, and the empirical illustration. However, it makes two assumptions: that there are no alternative explanations and that the current global trends will continue (Fig. 9.1; Box E). The case studies are examples of qualitative research, which provides rich information about the mechanisms responsible for the outcome (cf. Sect. 5.3: Causation in Qualitative Studies).

9.7.3 Study 3: Exploring the Importance of Social Processes for the Collapse of the Baltic Cod Stocks: A Modelling Approach

The goal of Lade et al. (2015) was to assess the role of social processes, such as fishers' decision making and actions, government decisions and market dynamics, for the collapse of the Eastern Baltic cod populations in the 1980s (Fig. 9.1; Box 3). The focus of the causal inquiry was thus to identify the causes of specified effects (Fig. 9.1; Box A.iii). The authors made sense of the causal configuration that may underlie the cod collapse through a collaborative process where the authors brought different ecological, economic and social-scientific expertise about Baltic cod fisheries to the discussions (Fig. 9.1; Boxes 1 and 2).

Together, they built a causal loop diagram specifying key feedbacks (cf. Sect. 5.4) that were hypothesised to have influenced the cod collapse (Fig. 9.1; Box B). Based on this diagram a generalised dynamical systems model was developed and parameterised for a situation before and during the beginning of the collapse, using fishery data, literature and expert knowledge of the research team (Fig. 9.11; Box C). A stability analysis of the modelled system separated in the social part, the ecological part and the coupled system before and during the collapse was conducted. This was done to assess the impact of the social system on the collapse and identify which feedbacks had the largest effect (Fig. 9.1; Box C).

The authors compared model versions, where the social and the ecological systems were decoupled, with a version of a coupled system, both before and after the collapse, to evaluate the causal influence of social processes on the collapse of the cod stocks. This is an example of the use of counterfactual reasoning (cf. Chap. 4) within a model, or rather different model versions, that represent the counterfactual situation. Based on a comparison, the authors developed causal knowledge about which social processes contributed to the shift in the Baltic Sea ecosystem (Fig. 9.1; Box D). Using the an analysis of the feedback mechanisms, (cf. Sect. 8.5.2: Feedback mechanisms), the authors explained the model outcomes. They made the causal claim that the adaptivity of external fishers (i.e. fishers that came to the Baltic Sea from Sweden's West Coast) initially stabilised the ecosystem despite changing environmental conditions for a certain period of time.

9.7.4 Study 4: Explaining Emergent Patterns of Self-Governance Arrangements in Small-Scale Fisheries: A Modelling Approach

Lindkvist et al. (2017) made a modelling study aimed to investigate the conditions under which either cooperative or non-cooperative forms of self-governance emerge in a typical fishing community in northwestern Mexico (Fig. 9.1; Box 3). It asks two causal questions; what are the causes of specified effects and how are they brought about (Fig. 9.1; Box A.i–ii)?

The study drew on frameworks and theories such as the SES framework (Ostrom, 2007), institutional analysis, collective action theory, common pool resource theory, and complex adaptive systems theory, all of which the researchers in the team previously had used (Fig. 9.1; Box 1, 2). Thus, the researchers' backgrounds influenced the study through their previous engagement with these theories and frameworks, but also through prior knowledge of the case and their experiences of working with fishers and in fishing communities (over 20 years for one co-author). This informed how they defined the social-ecological system and phenomena of interest, the research goals, and the assumptions of which variables matter for cooperatives such as the form of self-governance to persist over time (Fig. 9.1; Box B).

Against this background and based on data collected in previous studies the authors built a model that can be used as a virtual laboratory to answer the following research questions: (i) How do micro-level factors related to trust—such as the reliability of fishers, and loyalty between fish buyers and fishers and between members in cooperatives—affect the emergence and persistence of different self-governance arrangements? (ii) How does environmental variability affect whether cooperatives or patron-client relationships emerge as the dominant form of self-governance? (iii) How stable are these two self-governance arrangements and what causes them to fail?

Using the case knowledge and an agent-based model, the authors were able to discover and reason about specific mechanisms that explain how effects were produced. In the model one can change different variables and observe their effects on the emergence and persistence of different self-governance arrangements. This indicates which variables one can manipulate through different policies or interventions in relation to the desired outcome in reality. (cf. Sect. 6.7: Causation, Manipulation and Intervention). Additionally, the model setup includes several feedback mechanisms at the level of individual agents, such as the reinforcing feedback loop where increased loyalty results in less cheating, which in turn increases loyalty (cf. Sect. 8.5.2: Feedback mechanisms). These feedbacks became important parts of the explanation why under some conditions cooperatives could survive while not in others.

The model showed that high diversity in fishers' reliability and low initial trust between cooperative members make the establishment of cooperatives difficult. In contrast, patron-client relationships are more flexible in choosing whom to work

with and can better cope with this kind of diversity. However, once established, cooperatives are better equipped to handle seasonal variability in fish abundance and provide long-term security for the fishers. Through these types of causal findings gained from analysing and testing the model, combined with case based knowledge, the researchers could uncover and reason about specific causal mechanisms that help explain how certain effects are produced (cf. Sect. 8.4: Causal Explanations and Mechanisms).

The primary aim of the model was to investigate under which conditions different causes would, or would not, lead to certain effects and provide causal explanations for why and how (Fig. 9.1; Box A.ii). The causes and effects of interest to explore in the first place were, however, derived from prior case knowledge, theories and frameworks (Fig. 9.1; Box 2). The study design and the choice of agent-based modelling as a method (Fig. 9.1; Box C), were also a result of previous experience of the researchers involved (Fig. 9.1; Box 1–2). The interpretation of the results and the causal claims made (Fig. 9.1; Boxes D, E) were based on the key factors and processes in the model, but situated against background knowledge of the author team. The knowledge about the causal configuration of the social-ecological fishery system contributed to deeper knowledge about how policies could support specific governance structures in theory and practice (Fig. 9.1; Box E, 1, 2).

9.7.5 Study 5: Addressing the Challenges of Climate Adaptation: A Practice-Based Approach to Transdisciplinary Sustainability Interventions

The goal of the transdisciplinary 'Future proofing Conservation project' (van Kerkhoff et al., 2019) was to develop new ways of addressing the challenges posed by climate adaptation for protected area policy-makers and managers in Colombia. A practice-based approach to sustainability interventions is compatible with the assumption that for many sustainability problems 'optimal' solutions hardly exist and that problem formulations often are unclear and contested. It thus challenges linear assumptions about knowledge and action and suggests that 'the primary task of participants in sustainability interventions is to arrive at workable solutions to situations of dynamic complexity that are fundamentally open-ended and unpredictable' (West et al., 2019).

Accordingly, this example emphasises that causal reasoning in the context of complex SES requires collaborative and participatory processes involving the stake-holders affected in a particular place. Strictly speaking, collaboration is not only required when defining the causal configuration (Fig. 9.1; Box B), but already when characterising the prior causal understanding of the system (Fig. 9.1; Box 1, 2). In the process, climate science acquires a new role, 'not as a solution-provider ("let's wait until the scientists tell us what to do"), but as a knowledge base that conservation governance practitioners need to act upon ("we are knowledgeable actors")' (op.cit.,

547–8). Accordingly, the study encourages stakeholders to conceive of adaptation not as a simple 'once and for all' application of knowledge but as a continuously evolving practice. It thus highlights the importance of the feedback from the study itself to a continuously evolving causal understanding (Fig. 9.1; Box1, 2). There is a strong emphasis of practice-based approaches on this last point: Acting and knowing are merged in practice and as such '....the final methodology can be regarded as encouraging practitioners to think about climate adaptation as a practice, rather than a task. As a practice it is ongoing, deliberative and potentially transformative, framed by learning and dialogue rather than the application of technical solutions' (op.cit., 548).

9.8   Summary of Examples

These examples show that causal reasoning can be done in many different ways and it is strongly influenced by the goal of the study, by who is involved and what theories and frameworks, literature, scientific norms, and experiences, they bring to the table. For example, the first study builds on the potential outcomes framework (cf. Chap. 4) and research in economics in order to quantify the effect of an intervention using an experimental design that compares treatment with control units. The last study builds on practice theory and research in the humanities in order to build causal understanding through collaborative processes where scientists and non-scientists make sense of causality while engaging with the complexity of the problem and potential solutions. Here causal understanding is dynamic and continuously co-produced through the practice of problem solving. The examples illustrate the use of the causal concepts introduced earlier in the book, but also show that causal reasoning in SES research makes use of a broader set of concepts than what we could discuss in this introductory text.

Differences of causal reasoning and resulting causal claims between studies may arise because of different foci, e.g. on singular versus general causation (cf. Sect. 5.2), different data, e.g. quantitative versus qualitative, or different goals, e.g. evaluating the magnitude of a causal effect versus developing causal explanations. Study 1 for example makes use of the causal ideas of intervention and potential outcomes using an experimental design to collect quantitative data on a population of cases, study 2 applies INUS conditions and Hill's criteria using qualitative data. Studies 3 and 4 use counterfactual reasoning and manipulation in the context of modelling with the aim to identify mechanisms that bring about the phenomena of interest in the modelled system. The last study illustrates a focus of causal inquiry that lies specifically on how the causal configuration constitutes and re-constitutes in processes of transdisciplinary collaboration over the practice of climate change adaptation. So doing, this approach goes beyond the distinction made between causal and constitutive explanations made in Chap. 8 and explores how these interrelate and condition each other.

It is important to realise that the goals of a causal inquiry and choices made early on in a study direct causal reasoning and create path dependencies. For example, taking a systemic view and choosing a modelling approach will shape the process of making sense of the causal configuration differently than if a researcher takes a practice-theoretical view choosing a participatory approach. The five examples show that all elements of the causal reasoning processes are considered, but each study has a different emphasis. For example, in study 1, on the effectiveness of community monitoring, the authors put most emphasis on justifying their causal claims through scrutinising the design of their experiment and finding evidence, for the proposed mechanism that links the intervention to the outcomes. In study 4, on the emergence of self-governance, the focus is on understanding the conditions and the mechanism that explain why cooperatives rarely dominate. In study 5, on climate change adaptation, the authors emphasise the collaborative, practice-based and continuously evolving nature of causal reasoning. Study 3 puts much emphasis on building a comprehensive representation of the causal configuration through integrating interdisciplinary expert knowledge. Study 2 puts much emphasis on constructing an archetypical representation of the problem that is then tested in two case studies.

The five examples show that all elements of the causal reasoning processes are considered, but each study has a different emphasis. The five examples employ different methods for their causal inquiry, from experiments (study 1), dynamical systems and agent-based modelling (studies 3 and 4), participatory processes (study 5) and a combination of theoretical deliberations and case studies (study 2). These methods not only allow them to do different things, e.g., only the first method allows quantifying a causal effect, or only the modelling methods allow investigating how the system changes over time as a consequence of interactions between agents and their environment or feedbacks between system elements. They are also grounded in different assumptions of what is considered appropriate evidence for a causal claim. Finally, approaches and associated methods differ in their assessment of the causal configuration, from a focus on a single cause-effect relationship embedded in a larger causal configuration to a systemic view that incorporates more aspects of the larger causal configuration. The degree of formalisation of the causal configuration also varies, which has effects on which methods can be applied.

Our description of the studies also illustrate the difficulties of eliciting information about the worldviews, positionalities, and experiences that underlie causal reasoning because they are rarely made explicit. This lack of transparency is problematic because it limits our ability to assess the scope, quality and compatibility of a causal claim, or an approach to studying causation for a particular problem at hand, or for the integration of approaches or knowledge.

Tools that support eliciting underlying ontological and epistemological assumptions e.g., (Eigenbrode et al., 2007; Hazard et al., 2019)), and, more specifically, tools that facilitate dissecting the causal reasoning of different approaches help to increase transparency (we will briefly introduce this tool below). This is important in order to enable inter—and transdisciplinary collaborations across different traditions of causal reasoning.

9.9   Navigating the Diversity of Causal Reasoning

In this chapter, we have discussed and illustrated that causal reasoning takes place during all phases of a research process and that it is diverse and depends on the backgrounds, positionalities and prior knowledge of those involved in the study (Fig. 9.1). Awareness of the many steps in which causal reasoning manifests itself enables researchers to articulate, understand and reflect on the causal reasoning that underlies a particular study. This is important for inter- and transdisciplinary collaboration and for assessing the scope and validity of a causal claim, and its consequences for intervening in a SES to bring about a desired effect.

In order to make explicit how, specifically, the background, theories, research goals, etc. (everything in Boxes 1–3) influence causal reasoning activities during the different research phases, we have developed a guide called CoMap (Hertz et al., 2024). CoMap specifies five elements that together constitute causal reasoning: the conceptualisation of causation a study builds on, its analytical focus, the theories and frameworks used, the selected methods and causal notions. These elements are interdependent and influence each other, and their interplay is shaped by the purpose of an analysis. In addition, typically, one of these elements—which may be called 'entry point'—is particularly important in that it designates an element that orients or exerts influence on the other elements in that these need to 'align' with it.

Through making these choices and path dependencies explicit, this guide reveals how causal reasoning looks like, that is, it becomes apparent which choices can and need to be made by researchers in assembling a study. The examples presented above show how, accordingly, causal reasoning might vary considerably. This means that through the process of eliciting causal reasoning we become aware of each other's 'blind spots'. That provides the basis for (1) a reflection on the assumptions underlying the causal reasoning of research approaches, for (2) engaging in inter- and transdisciplinary collaboration, either by developing a common research approach, or (3) by relating different research approaches to each other.

9.10   Summary

This chapter characterise causal reasoning as the cognitive activities we engage in, implicitly or explicitly, when studying relations between causes and effects. We engage in causal reasoning during the entire research process of a study, from developing the causal questions, making sense of the causal configuration of the phenomenon of interest, designing the study and choosing appropriate methods, interpreting results to make and justify causal claims. In our causal reasoning we draw on our theoretical and methodological background, including ideas about causation, on previous experiences, the literature, the norms of our community and our previous understanding of the SES in which the phenomenon of interest

is embedded. Differences in the goals of a causal inquiry, such as prediction, explanation, or intervention, and in the background conditions that shape causal reasoning, produce a large diversity of causal reasoning strategies. These different strategies produce different causal understanding, which has consequences for what can be done with the knowledge, e.g., when designing interventions.

Further Reading

- Schlüter et al. (2023a)
- Hertz et al. (2024)
- Peña (2023)
- Peña and Ylikoski (2023)
- Banitz et al. (2022b)
- Banitz et al. (2022a)

Study Questions

- Have you encountered different ways of causal reasoning? If so, which ones?
- How does your worldview, positionality, and experiences influence your causal reasoning?
- How has the literature, theories, and frameworks you have engaged with influenced your causal reasoning?
- How have the scientific norms at your home institution hindered or enabled your ability to engage in different types of causal questions?
- Can you think of examples from your own experience where your understanding of the causal configuration changed after you conducted a study and observed a particular phenomenon? Did this experience change your worldview and/or what literature, theories, and frameworks you engage with?
- Do you see any other interesting differences or similarities between the examples presented in Sect. 9.4?

References

Banitz T, Schlüter M, Lindkvist E, Radosavljevic S, Johansson L-G, Ylikoski P, Martínez-Peña R, Grimm V (2022a) Model-derived causal explanations are inherently constrained by hidden assumptions and context: the example of baltic cod dynamics. Environ Model Softw 156(Oct):105489

Banitz T, Hertz T, Johansson L-G, Lindkvist E, Martínez-Peña R, Radosavljevic S, Schlüter M, Wennberg K, Ylikoski P, Grimm V (2022b) Visualization of causation in social-ecological systems. Ecol Soc 27(1).

Biggs R, Clements H, de Vos A, Folke C, Manyani A, Maciejewski K, Martín-López B, Preiser R, Selomane O, Schlüter M (2021) What are social-ecological systems and social-ecological systems research? In: The Routledge handbook of research methods for social-ecological systems. Routledge, London

Biggs R, De Vos A, Preiser R, Clements H, Maciejewski K, Schlüter M (2022) The Routledge handbook of research methods for social-ecological systems. Routledge/Taylor & Francis Group, Abingdon/Oxon

Caniglia G, Schlüter M (2023) Practical causal knowledge for sustainability: Implications of co-production for a philosophical understanding of causality in sustainability science. https://doi.org/10.31235/osf.io/cg9nk

Carpio MBD, Alpizar F, Ferraro PJ (2021) Community-based monitoring to facilitate water management by local institutions in costa rica. Proc Natl Acad Sci 118(29)

Chambers JM, Wyborn C, Ryan ME, Reid RS, Riechers M, Serban A, Bennett NJ, Cvitanovic C, Fernández-Giménez ME, Galvin KA, Goldstein BE, Klenk NL, Tengö M, Brennan R, Cockburn JJ, Hill R, Munera C, Nel JL, Österblom H, Bednarek AT, Bennett EM, Brandeis A, Charli-Joseph L, Chatterton P, Curran K, Dumrongrojwatthana P, Durán AP, Fada SJ, Gerber J-D, Green JMH, Guerrero AM, Haller T, Horcea-Milcu A-I, Leimona B, Montana J, Rondeau R, Spierenburg M, Steyaert P, Zaehringer JG, Gruby R, Hutton J, Pickering T (2021) Six modes of co-production for sustainability. Nat Sustain 4(11):983–996

Eigenbrode SD, Orourke M, Wulfhorst JD, Althoff DM, Goldberg CS, Merrill K, Morse W, Nielsen-Pincus M, Stephens J, Winowiecki L, Bosque-Pérez NA (2007) Employing philosophical dialogue in collaborative science. BioScience 57(1):55–64

Geels FW (2022) Causality and explanation in socio-technical transitions research: mobilising epistemological insights from the wider social sciences. Res Policy 51(6):104537

Hazard L, Cerf M, Lamine C, Magda D, Steyaert P (2019) A tool for reflecting on research stances to support sustainability transitions. Nat Sustain 3(2):89–95

Hertz T, Banitz T, Martinez-Peña R, Radosavljevic S, Lindkvist E, Johansson L-G, Ylikoski P, Schlüter M (2024) Eliciting the plurality of causal reasoning in social-ecological systems research. Ecol Soc 29(1)(14)

Homer-Dixon T, Walker B, Biggs R, Crépin A-S, Folke C, Lambin EF, Peterson GD, Rockström J, Scheffer M, Steffen W, Troell M (2015) Synchronous failure: the emerging causal architecture of global crisis. Ecol Soc 20(3)

Illari P, Russo F (2014) Causality: philosophical theory meets scientific practice. Oxford University Press, Oxford

Kates RW (2011) What kind of a science is sustainability science? Proc Natl Acad Sci 108(49):19449–19450

Kimmel K, Dee LE, Avolio ML, Ferraro PJ (2021) Causal assumptions and causal inference in ecological experiments. Trends Ecol Evol 36(12):1141–1152

Lade SJ, Niiranen S (2017) Generalized modeling of empirical social-ecological systems. Nat Resour Model 30(3)

Lade SJ, Niiranen S, Hentati-Sundberg J, Blenckner T, Boonstra WJ, Orach K, Quaas MF, Österblom H, Schlüter M (2015) An empirical model of the baltic sea reveals the importance of social dynamics for ecological regime shifts. Proc Natl Acad Sci 112(35):11120–11125. https://doi.org/10.1073/pnas.1504954112

Lang DJ, Wiek A, Bergmann M, Stauffacher M, Martens P, Moll P, Swilling M, Thomas CJ (2012) Transdisciplinary research in sustainability science: practice, principles, and challenges. Sustain Sci 7(S1):25–43

Lindkvist E, Basurto X, Schlüter M (2017) Micro-level explanations for emergent patterns of self-governance arrangements in small-scale fisheries—a modeling approach. PLOS ONE 12(4):e0175532

Norström AV, Cvitanovic C, Löf MF, West S, Wyborn C, Balvanera P, Bednarek AT, Bennett EM, Biggs R, de Bremond A, Campbell BM, Canadell JG, Carpenter SR, Folke C, Fulton EA, Gaffney O, Gelcich S, Jouffray J-B, Leach M, Tissier ML, Martín-López B, Louder E, Loutre M-F, Meadow AM, Nagendra H, Payne D, Peterson GD, Reyers B, Scholes R, Speranza CI, Spierenburg M, Stafford-Smith M, Tengö M, van der Hel S, van Putten I, Österblom H (2020) Principles for knowledge co-production in sustainability research. Nat Sustain 3(3):182–190

Ostrom E (2007) A diagnostic approach for going beyond panaceas. Proc Natl Acad Sci 104(39):15181–15187

Peña RM, Schlüter M, Hertz T, Banitz T, Wennberg K, Radosavljevic S, Ylikoski P (2023) Analysis of causal argumentation in social-ecological systems research. https://doi.org/10.31235/osf.io/q9hga

Peña RM, Ylikoski P (2023) Constructing the Coleman boat – Mechanism-based theorising in socio-ecological research. https://doi.org/10.31235/osf.io/wnrpz

Preiser R, Schlüter M, Biggs R, García MM, Haider J, Hertz T, Klein L (2021) Complexity-based social-ecological systems research: philosophical foundations and practical implications. The Routledge handbook of research methods for social-ecological systems. Routledge, London, pp 27–46

Rubin DB (2005) Causal inference using potential outcomes: design, modelling, decisions. J Am Stat Assoc 100(469):322–331

Rubin DR (2004) Direct and indirect causal effects via potential outcomes (with discussion). Scand J Stat 31:161–179, 195–198

Schlüter M, Hertz T, García MM, Banitz T, Grimm V, Johansson L-G, Lindkvist E, Rodrigo M-P, Radosavljevic S, Wennberg K, Ylikoski P, Ylikoski P (2023a) Navigating causal reasoning in sustainability science. https://doi.org/10.31235/osf.io/kn49v

Schlüter M, Brelsford C, Ferraro PJ, Orach K, Qiu M, Smith MD (2023b) Unraveling complex causal processes that affect sustainability requires more integration between empirical and modeling approaches. Proc Natl Acad Sci 120(41)

van Kerkhoff L, Munera C, Dudley N, Guevara O, Wyborn C, Figueroa C, Dunlop M, Hoyos MA, Laura Becerra JC (2019) Towards future-oriented conservation: managing protected areas in an era of climate change. Ambio 48(7):699–713

West S, van Kerkhoff L, Wagenaar H (2019) Beyond linking knowledge and action: towards a practice-based approach to transdisciplinary sustainability interventions. Policy Stud 40(5):534–555

Appendix A
How Does a Theory Relate to Reality?

In order to be of any practical use a theory must say something about some part of reality. A theory consists of sentences, which in turn consists of words. So how do words and sentences relate to reality? How do they get meaning?

We here focus on those words used as singular terms. This is the crucial point, because if we have established that the noun phrase (which is a singular term) in a sentence refers to something in the real word, we have a sentence that is *about* something. How is this reference relation established?

It is obvious that an object can be given any randomly chosen name. Hence, in order for a name to be coupled to a particular thing, there must be something that connects the name to the object. Could that be fixed by a definite description, such as 'The tallest man in town' or 'the capital of France'?

When we say, e.g., that Paris is the capital of France, the referent of the term 'Paris' is identified by the description 'The capital of France'. But this only postpones the question, for now we ask for the meanings of 'capital' and 'France'. We end up with either an endless regress, or else a circularity.

Obviously, there must be and endpoint somewhere, and the regress ends where we are able to identify the referent of a term without using other words. This is reached when one can point to the object in question and show an audience that *one's use* of a certain name now means the object pointed to. Thus the basic contact points between language and reality consist of expressions used together with pointing gestures, (such expressions are in linguistics called 'deixis') aimed at identifying the referent of a name, pronoun or description. These gestures are *actions* performed by language users and observed by other language users. In short, reference relations are ultimately established by extra-linguistic activities. Without such a basis no connection between linguistic items and things, events or states of affairs in the real world is possible.

This is not only an observed feature of ordinary language use but, more profoundly, a consequence of Löwenheim-Skolem's theorem, as is shown in (Johansson, 2021). This theorem of mathematical logic says roughly that if a theory

T has a consistent interpretation in an infinite domain of objects, i.e, has a true model in such a domain, it is also true in a model consisting of the natural numbers.[1] It follows immediately that no consistent theory can determine what it is about; it can always be interpreted as being a theory about the natural numbers. If we, in order to single out the intended interpretation, add interpretation rules to a theory T, we get another theory T'. But Löwenheim-Skolem's theorem also applies to T'. So the connection between a theory T and what T is about must be established by non-theoretical means, in communication contexts. That is done ultimately by use of demonstratives and gestures.

This may work when we talk about observable objects. But how do we refer to unobservable objects? They cannot as matter of principle be identified by gestures, ostensive definitions.

We may distinguish between two cases, concrete and abstract objects. By a concrete object we mean any object that exists in space and time, observable or not. An abstract object, by contrast does not exist in space and time. Typical abstract objects are properties, relations, sets and numbers.

Atoms and molecules are concrete objects and they were for a long time believed to be unobservable. However, observability depends on technology; nowadays one can, using advanced technology, in certain cases observe individual atoms. But even before such technology was available, one could identify and refer to atoms, at least in some situations, by indirect means. Observing a state change of a measuring device, one may, using theory, say e.g. 'The atom hitting the measuring device at time t_0 had an energy of 3 eV.'. In this sentence the singular term 'the atom hitting the measuring device at t_0' refers to a concrete but not directly observed thing.

Referring to abstract objects is a bit more complicated. Let us concentrate on sets, collections of objects.

A set is an abstract entity, even if its members are concrete things. But we can identify a particular set by identifying its members. So if the members of a set are concrete objects, one can identify the set indirectly by pointing to its members, saying something like 'these things together make up the set S'. And one can continue by constructing sets of sets, sets of sets of sets, etc. What is needed is a basic level of concrete objects.

Many abstract things can be defined in terms of sets (for example numbers) but what about those that cannot? Well, it is doubtful if talk about such things has any clear meaning. This is a highly controversial topic in philosophy.

The fact that the reference relation must be based on extra-linguistic activities together with context dependent expressions, demonstratives, is often overlooked by many theoreticians in all disciplines. The reason is two-fold: (i) many theoretical expressions used in a theory are used in ordinary language, and (ii) the referents

---

[1] One may observe that in logic and *formal* semantics one uses the word 'model' in another sense than in empirical sciences. In logic and metamathematics one conceives of theories as structures of symbols without any interpretation at all. A model of such a theory is a consistent attribution of the truth-value 'true' to the sentences in the theory. In empirical sciences, by contrast, a model is a simplified description of real objects, phenomena and states of affairs.

of many terms in ordinary discourse have been settled by such extra-linguistic interactions involving at least some people and is usually tacitly presupposed by theoretical scientists.

The important conclusion is that no description, however detailed it is, is by itself sufficient for identifying what is described. Concrete interactions between humans, involving gestures and use of demonstratives, are needed for the identification of some of the objects talked about. It follows that a pure theory, lacking any connection or description of observable things and events, cannot say anything whatsoever about the empirical world.

This has profound consequences for discourse about desires, beliefs, intentions and other mental states, since these entities cannot be directly identified by gestures and demonstratives. Hence, any theory, in which such entities are postulated as the causes of events, must be supported by evidence obtained by reports about observable events.

Appendix B
Models

Social-ecological systems consist of many parts, agents, mechanisms, etc. These parts act and react to the states and state changes of other parts of the system: social-ecological systems are complex adaptive systems. The problem for policy-makers is to obtain knowledge about these adaptive mechanisms, sufficient for effective interventions and policy decisions. Here are two quotes describing the challenge:

> While economic theory has often successfully ignored most complexity in modelling economic systems, research on social-ecological systems shows that it can be very misleading to do so. Complexity entails substantial modelling challenges, but simple models can incorporate some elements of complexity to provide novel insights. Dynamical systems are starting points for modelling social-ecological systems, and agent-based models provide a natural extension that better incorporates heterogeneity among individuals. (Levin et al., 2012, sec. 3)

> In economic systems and ecological systems alike, heterogeneity introduces complexities of essential importance, motivating efforts to model these features. Agent-based models or individual-based models allow each individual to have unique behaviors that may change in response to others' actions, and the possibly slow evolution of macroscopic variables (Bonabeau, 2002; Couzin et al., 2005; Grimm et al., 2005). These models easily implement detailed assumptions about individual behavior, but suffer from a lack of analytic tractability and difficulties with extracting robust conclusions. Thus, it is important that these descriptions ultimately be embedded into an analytical framework that helps to understand the statistical mechanics of these heterogeneous ensembles (Flierl et al., 1999; Couzin et al., 2011). (Levin et al., 2012, sec. 3.5)

One can never give a complete description of a complex system; one is forced to construct models, simplified descriptions, which leave out many details but hopefully take into account the most salient parts and their most relevant interactions.

One may think that models of SES that include more factors give better fit with observations and increased predictive power than those including fewer ones. On the other hand, a model with many variables may not be useful when deciding

which interventions to perform in order to attain a certain goal. This tension was commented on by García-Callejas and Araújo (2016):

> How complex does a model need to be to provide useful predictions is a matter of continuous debate across environmental sciences. In the species distributions modelling literature, studies have demonstrated that more complex models tend to provide better fits. However, studies have also shown that predictive performance does not always increase with complexity. Testing of species distributions models is challenging because independent data for testing are often lacking, but a more general problem is that model complexity has never been formally described in such studies. Here, we systematically examine predictive performance of models against data and models of varying complexity. (García-Callejas and Araújo, 2016, 4)

One and the same system can be modelled in several different ways, depending on what kind of inferences one want to be able to draw. Quite often in SES research the goal is to construct a model enabling us to understand what to do in order arrive at sustainable use of a natural resource. We have discussed this in more detail in (Banitz et al., 2022a).

When constructing models of complex systems it is advisable to do a cost-benefit analysis; the cost in terms of time and coding effort of constructing a more complex model may not correlate to any clear increase in predictive efficacy, as discussed in (Grimm et al., 2005). Figure B.1 is adapted from that paper.

These reflections tell us that it is not the complexity of a system, object or state of affairs in itself that is of relevance; the question concerns the complexity of our *models* in the discussion about complex causation. This point was made by Allen et al. (2018):

> Much discussion of complexity is confused because complexity is mistaken as a material issue. Complexity arises from the way the situation is addressed, and is not material in itself. (p. 39)

So complexity is to be understood as an attribute of models or descriptions of SES systems. And the first question is; are there any measure of complexity, so that one can compare models in terms of degree of complexity? And how is complexity related to predictability? Czeslaw Mesjasz touched on this topic (Mesjasz, 2010, 708):

> In order to identify the meaning of complexity, based on some properties of the relationships between human observers, or observing systems in general, and all kinds of observed systems, natural and artificial, including the social ones, Biggiero (2001, 3, 6) treats predictability of behaviour of an entity as the fundamental criterion for distinguishing various kinds of complexity. As a foundation he proposes an interpretation of complexity as a property of objects which are neither deterministically nor stochastically predictable. "Complexity" refers to objects which are predictable only in a short run and that can be faced only with heuristic and not optimising strategies. (Biggiero, 2001, 6)

The last sentence expresses the crucial idea that complexity entails unpredictability. But this formulation is not the best one, it would be more useful to explicate both complexity and predictability as measures, assuming that the more complex a system is, the shorter time span during which we can make useful predictions.
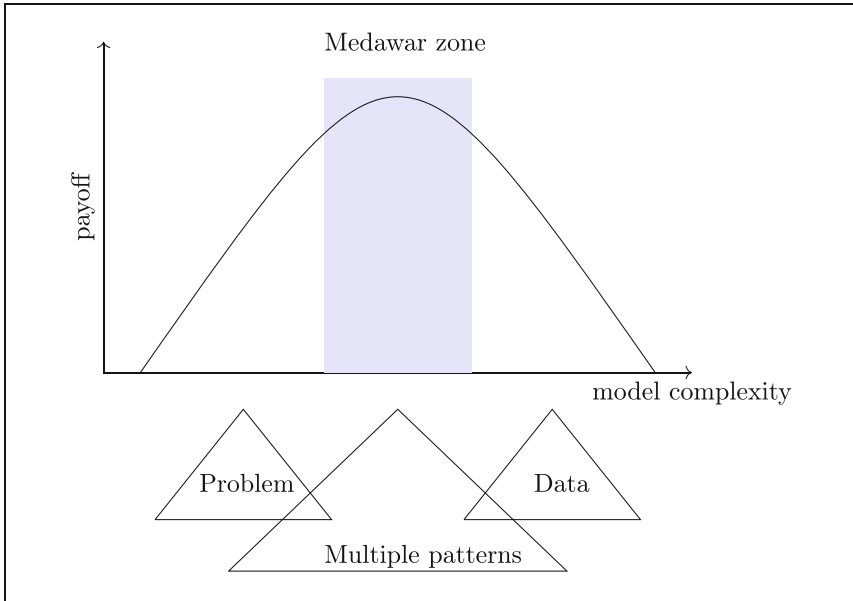
Fig. B.1 Payoff of bottom-up models versus their complexity. A model's payoff is determined not only by how useful it is for the problem it was developed for, but also by its structural realism; i.e., its ability to produce independent predictions that match observations. If model design is guided only by the problem to be addressed (which often is the explanation of a single pattern), the model will be too simple. If model design is driven by all the data available, the model will be too complex. But there is a zone of intermediate complexity where the payoff is high. We call this the 'Medawar zone' because Medawar described a similar relation between the difficulty of a scientific problem and its payoff (Loehle, 1990). If the very process of model development is guided by multiple patterns observed at different scales and hierarchical levels, the model is likely to end up in the Medawar zone. Adapted from (Grimm et al., 2005, 988)

Is there any suggested measure of complexity in the literature? Yes. The most developed idea about degree of complexity, which also is of optimal generality, is *algorithmic complexity*, first expressed by Solomonoff (1964a,b) and Kolmogorov (1998/1963) and developed by Chaitin (1987). (It goes under several names, Kolmogorov complexity, Kolmogorov-Chaitin complexity or algorithmic complexity. The general idea is that the degree of complexity of a system is measured by the shortest algorithm which can produce a description of that system. The interested reader is referred to these papers.

Appendix C
Confidence Intervals and Correlations

C.1 Confidence Intervals

Empirical data are very often drawn from samples of a population of some kind, be it a collection of objects, persons, states of affairs, situations, events or whatever. It is obvious that the value of a parameter observed in the sample can be far from its value in the entire population. So a crucial task is to estimate the real value of the parameter, i.e., the value it has in the entire population.

The technique to do this is to calculate a confidence interval, which allows one to conclude that the parameter value with a certain probability lies within the calculated interval.

When calculating the confidence interval one must use a probability distribution function. How to chose?

If the sampling is randomised, one can justify the use of a normal distribution. The reason is that it is a remarkable fact that if we randomly select a number of items from a population, the mean values of a chosen variable in a series of such samplings will be approximately normally distributed around its mean in the entire population. For example, the true mean of a series of rolling a dice is 3.5. A short series och such rollings will most often not give exactly the mean 3.5. But if one performs a great number of such series of dice rolling, the distribution of the means of these series will get closer and closer to the normal distribution centred at 3.5.

If sampling is not random one cannot say anything about the relation between observed values in the sample and the true values in the population. But if sampling is randomised, one can use a normal distribution for calculating the confidence interval for the true parameter value. The crucial step is to use the central limits theorem:

Central Limit Theorem Draw a simple random sample of size n from a population with mean μ and standard deviation σ. When n is large, the sample mean $\bar{x}$ is approximately normal distributed with mean μ and standard deviation $s = \sigma/\sqrt{n}$.

As a rule of thumb, 'large' means $n \geqslant 30$. Now we can calculate a confidence interval for the population mean μ.

Suppose we want to calculate an interval such that μ with 95% probability is inside this interval. That means that we shall calculate an interval of $\bar{x} \pm 1.96\sigma$[1] In other words, a confidence interval of $\bar{x} \pm 1,96\sigma$ contains with 95% probability the real mean.

The figure below shows a normal distribution function with $\mu = 20$ and $\sigma = 2$. The total area under the curve is 1 and represents the total probability. The area under the curve between 16.08 and 23.92 (i.e., $20 \pm 1.96\sigma$) is 95% of the total area.

This allows us to say that the sample mean $\bar{x}$ with 95% probability lies between 16.08 and 23.92. Since we in fact know $\bar{x}$, we can infer that the real mean μ with 95% probability is in the interval [16.08, 23.92]

$$\mathrm{N}(20,2) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2 \cdot 2^2}(x-20)^2}$$
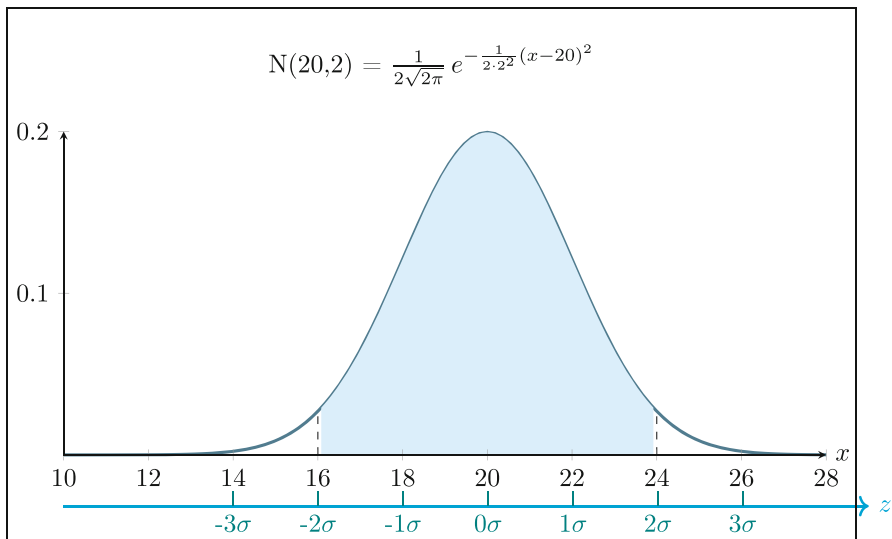
So the method is as follows:

1. Draw a random sample of n items from the population.
2. Calculate the sample mean $\bar{x}$, the sample standard deviation s and the standard deviation σ in the population.
3. Chose a desired probability and calculated the confidence interval.

---

[1] A 95% confidence interval is such that it is 2.5% probability that the real value is above the interval, and likewise 2.5% below the interval. Using the function NORM.INV in Excel, one can calculate the width of an interval, given a desired probability. NORM.INV takes three inputs; the mean, standard deviation and required probability. It returns the inverse to the normal distribution, i.e. the probability that the real value is below the returned value. So for a 95% confidence interval you should put in 0.025+0.95=0.975 as the third input, which, with mean =0 and std=1 returns 1.96.

C.2   Confidence Intervals for Correlations

There is a complication when calculating the confidence interval for a correlation. It is due to the fact that by definition the coefficient of correlation is such that $-1 \leqslant r \leqslant 1$, which means that the probability distribution must be skewed if r is near -1 or 1. Suppose, for example, we have observed in a sample a correlation r=0.80 and calculated the standard deviation to be 0.15. The normal distribution N(0.80, 0.15) can be seen below.

$$\mathrm{N}(0.80, 0.15) = \frac{1}{0.15\sqrt{2\pi}}\, e^{-\frac{1}{2\cdot 0.15^2}(x-.80)^2}$$
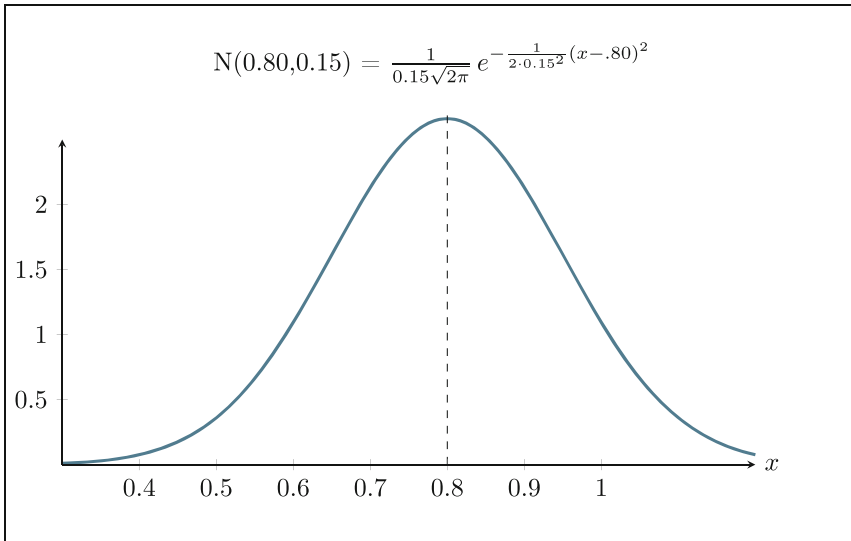
This cannot represent the real situation because the coefficient of correlation is always less than 1. In other words, the real probability distribution must be skewed towards the left. (Similarly, if the correlation is near zero, the probability distribution must be skewed to the right.) What to do?

Fisher (1915) solved the problem by performing a transformation, called Fisher's z-transformation

$$z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) = \operatorname{artanh}(r), \tag{C.1}$$

which results in a nearly normal distribution. He also showed that the standard deviation is

$$\sigma = \frac{1}{\sqrt{N-3}}, \tag{C.2}$$

Now we can use this z for calculating a confidence interval. But no cumbersome calculations are in practice needed, the discussion above is only meant as expla-

nation. One can use e.g., https://www.statskingdom.com/correlation-confidence-interval-calculator.html, where by plugging in the observed coefficient of correlation in the sample, the sample size and the required confidence level the confidence interval is returned. In this case the 95% confidence interval becomes [0.69, 0.88]

Glossary

Association   The word 'association' has in scientific contexts a slightly more general meaning than in ordinary language. If two variables are correlated (the coefficient of correlation being substantially non-zero) they are said to be associated. But the converse is not true. Two variables may satisfy a non-linear relation, while the correlation coefficient may be zero. This would occur if e.g. a best fit curve is close to a second degree curve. (This is sometimes the form of dose-response curves in medicine.) It is furthermore not uncommon to interpret 'association' as expressing a causal relation, which is not correct. 89

Category   A category is a name for a set of objects of some kind given a common label, the category, such as 'gender' or 'cod'. Philosophers disagree on whether categories are things 'over and above' the individual items belonging to the category or not. 24

Category variable   A category variable is a mapping from a set of objects, events or observations to a category. 12, 102

Ceteris paribus   A condition of the form 'all else being the same'. 57

Coefficient of correlation   The coefficient of correlation between two variables X and Y is defined as

$$\rho_{XY} = \frac{E(X - E(X)) \cdot E(Y - E(Y))}{\sigma_X \sigma_Y}.$$ 62, 86

Conditional probability   The probability of an event in a subset of a population identified by a condition. 12, 49

Confounder   A variable that is correlated with two observed variables, a predictor and an outcome variable. When you control for a covariate and the correlation between predictor and outcome variable disappears, it was a confounder. 15, 101

Constitutive explanation   An explanation that consists of a description of the parts of a system, its constituents, and of how they interact. 4, 96

Counterfactual   A counterfactual is a conditional sentence where both the antecedent and consequent are known to be false, hence it is expressed in

subjunctive mood. Ex: If the British people had voted no to Brexit, UK would still be member of EU. 31

Covariate   Any variable that you may consider during your analysis in addition to your primary variables of interest. 86

Explanatory variable   A variable that is *believed* to be the cause of another variable called the response variable. 88

General statement   A statement of the form 'All x is F', where F is a description of a property. 35

General term   General terms can be *true of* several things, i.e. they stand in predicate position in complete sentences. 13

Intentional   Intentional concepts are used to describe mental states such as desires, beliefs and emotions, which all are in a certain sense 'directed' towards certain things, such as persons, things or events. The structure of an intentional expression is <act> directed towards <act-object>, for example '....happy for the present'. The directedness is indicated by a preposition, or in the case of contents of desires and beliefs with 'that'. 98

INUS-condition   A condition which is Insufficient but Necessary part of a complex of conditions which, as a complex, is Unnecessary but Sufficient. 51, 53

Joint probability   The probability of an event instantiating two types of events, e.g. the probability that a card drawn from a deck of cards is a club and an ace. 86, 87

Marginal probability   The probability of an event in an entire population. 12, 49

Natural experiment   A situation where by random, or in the course of events, there occurs a division of a population into two groups without intervention by the researcher and where the difference is described by a variable which might be a cause. 85

Quantitative variable   A quantitative variable is a mapping from a set of objects, events or observations to numbers. 12, 14, 18, 24

Response variable   A variable which is *believed* to represent the effect of another variable, called the explanatory variable; but there need not be any causal relation between the explanatory and response variable. 74, 88

SES   social-ecological system. 83

Singular term   A name or a definite description of at most one single thing. Singular terms can either be nouns or direct objects in complete sentences. A singular term may be empty, lacking reference. 13, 127

UGC   A Universally Generalised Conditional (UGC) is a sentence of the form 'For all x, if x is A, then x is B.'. 60, 62

Universal   A universal is that which many things may have in common, such as the colour blue, the property of being a tiger, or the relation of being father to. Whether universals really exist is a perennial debate in philosophy. 45

Bibliography

Allen TFH, Austin P, Giampietro M, Kovacic Z, Ramly E, Tainter J (2018) Mapping degrees of complexity, complicatedness, and emergent complexity. Ecol Complexity 35(Sep):39–44

Anderies JM, Rodriguez AA, Janssen MA, Cifdaloz O (2007) Panaceas, uncertainty, and the robust control framework in sustainability science. Proc Natl Acad Sci 104(39):15194–15199

Anderies JM, Mathias J-D, Janssen MA (2019) Knowledge infrastructure and safe operating spaces in social-ecological systems. Proc Natl Acad Sci 116(12):5277–5284

Anderies JM, Cumming GS, Clements HS, Lade SJ, Seppelt R, Chawla S, Müller B (2022) A framework for conceptualizing and modeling social-ecological systems for conservation research. Biol Conserv 275(Nov):109769

Angrist JD, Pischke J-S (2010) The credibility revolution in empirical economics: how better research design is taking the Con out of econometrics. J Econ Perspectives 24(2):3–30

Anscombe GEM (1971) Causality and determination : an inaugural lecture. Cambridge University Press, London

Banitz T, Schlüter M, Lindkvist E, Radosavljevic S, Johansson L-G, Ylikoski P, Martínez-Peña R, Grimm V (2022a) Model-derived causal explanations are inherently constrained by hidden assumptions and context: the example of Baltic Cod dynamics. Environ Model Softw 156(Oct):105489

Banitz T, Hertz T, Johansson L-G, Lindkvist E, Martínez-Peña R, Radosavljevic S, Schlüter M, Wennberg K, Ylikoski P, Grimm V (2022b) Visualization of causation in social-ecological systems. Ecol Soc 27(1)

Barbrook-Johnson P, Penn A (2022) Systems mapping: how to build and use causal models of systems. Palgrave Macmillan, London

Barnosky AD, Koch PL, Feranec RS, Wing SL, Shabel AB (2004) Assessing the causes of late pleistocene extinctions on the continents. Science 306(5693):70–75

Bechtel W, Abrahamsen A (2005) Explanation: a mechanist alternative. Stud Hist Philos Sci Part C: Stud Hist Philos Biol Biomed Sci 36(2):421–441

Becker HS (1986) Writing for social scientists. University of Chicago Press, Chicago

Biggiero L (2001) Sources of complexity in human systems. Nonlinear Dynamics Psychol Life Sci 5(3): 3–19

Biggs R, Clements H, Vos A de, Folke C, Manyani A, Maciejewski K, Martín-López B, Preiser R, Selomane O, Schlüter M (2021) What are social-ecological systems and social-ecological systems research? In: The Routledge handbook of research methods for social-ecological systems. Routledge, London

Biggs R, De Vos A, Preiser R, Clements H, Maciejewski K, Schlüter M (2022) The Routledge handbook of research methods for social-ecological systems. Routledge, Taylor & Francis Group, Abingdon, Oxon

Bonabeau E (2002) Agent-based modeling: methods and techniques for simulating human systems. Proc Natl Acad Sci 99(Suppl_3):7280–7287

Boyle R (1662) New experiments physico-mechanicall, touching the spring of air, and its effects. Printed by H. Hall for Tho. Robinson, Oxford

Brady U, Tellman B, Anderies JM (2015, November) Institutional Analysis of the Artisanal Green Turtle Fishery in Eastern (Caribbean) Nicaragua

Burhan NAS, Sabri MF, Rindermann H (2022) Cognitive ability and economic growth: how much happiness is optimal? Int Rev Econ 70:63–100

Caniglia G, Schlüter M (2023) Practical causal knowledge for sustainability: implications of co-production for a philosophical understanding of causality in sustainability science

Carlson J, Doyle JC (2002) Complexity and robustness. Proc Natl Acad Sci 99(Suppl 1):2538–2545

Carpio MBD, Alpizar F, Ferraro PJ (2021a) Community-based monitoring to facilitate water management by local institutions in Costa Rica. Proc Natl Acad Sci 118(29):e2015177118

Carpio MBD, Alpizar F, Ferraro PJ (2021b) Community-based monitoring to facilitate water management by local institutions in Costa Rica. Proc Natl Acad Sci 118(29)

Cartwright N (1983) How the laws of physics lie. Oxford University Press, Oxford

Cartwright N (1989) Nature's capacities and their measurement. Clarendon Press, Oxford

Cartwright N (2017) Can structural equations explain how mechanisms explain?, Chap. 8 In: Beebee H, Hitchcock C, Price H (eds) Making a difference: essays on the philosophy of causation. Oxford University Press, Oxford

Chaitin GJ (1987) Algorithmic information theory. Cambridge University Press, Cambridge

Chambers JM, Wyborn C, Ryan ME, Reid RS, Riechers M, Serban A, Bennett NJ, Cvitanovic C, Fernández-Giménez ME, Galvin KA, Goldstein BE, Klenk NL, Tengö M, Brennan R, Cockburn JJ, Hill R, Munera C, Nel JL, Österblom H, Bednarek AT, Bennett EM, Brandeis A, Charli-Joseph L, Chatterton P, Curran K, Dumrongrojwatthana P, Durán AP, Fada SJ, Gerber J-D, Green JMH, Guerrero AM, Haller T, Horcea-Milcu A-I, Leimona B, Montana J, Rondeau R, Spierenburg M, Steyaert P, Zaehringer JG, Gruby R, Hutton J, Pickering T (2021)0 Six modes of co-production for sustainability. Nat Sustain 4(11):983–996

Cliff N (1983) Some cautions concerning the application of causal modelling methods. Multivariate Behav Res 18:115–126

Collier JD (1999) Causation is the transfer of information. In: Sankey H (ed) Causation and laws of nature. Kluwer Academic Publishers, Dordrecht, pp. 215–246

Couzin ID, Krause J, Franks NR, Levin SA (2005) Effective leadership and decision-making in animal groups on the move. Nature 433(7025):513–516

Couzin ID, Ioannou CC, Demirel G, Gross T, Torney CJ, Hartnett A, Conradt L, Levin SA, Leonard NE (2011) Uninformed individuals promote democratic consensus in animal groups. Science 334(6062):1578–1580

Craver C (2007) Explaining the brain: mechanisms and the mosaic unity of neuroscience. Oxford University Press, Oxford

Csete ME, Doyle JC (2002) Reverse engineering of biological complexity. Science 295:1664–1669

Dawid AP (2000) Causal inference without counterfactuals. J Am Stat Assoc 95:407–448

Eberhardt F (2007) Causation and intervention. Ph.D. thesis, Carnegie Mellon University, Pittsburgh

Eberhardt F, Scheines R (2007) Interventions and causal inference. Philos Sci 74(5):981–995

Eells E, Sober E (1983) Probabilistic causality and the question of transitivity. Philos Sci 50:35–57

Eigenbrode SD, Orourke M, Wulfhorst JD, Althoff DM, Goldberg CS, Merrill K, Morse W, Nielsen-Pincus M, Stephens J, Winowiecki L, Bosque-Pérez NA (2007) Employing philosophical dialogue in collaborative science. BioScience 57(1):55–64

Elster J (1989) Nuts and bolts for the social sciences. Cambridge Univ. Press, Cambridge

Engemann K (2020) What is the Phillips curve (and why has it flattened)? https://www.stlouisfed.org/open-vault/2020/january/what-is-phillips-curve-why-flattened

Fisher RA (1915) Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika 10(4):507

Fisher RA (1925) Statistical methods for research workers. Oliver and Boyd, Edinburgh

Fisher RA (1951) The design of experiments, 6th edn. Oliver and Boyd, Edinburgh

Flierl G, Grünbaum D, Levins S, Olson D (1999) From individuals to aggregations: the interplay between behavior and physics. J Theor Biol 196(4):397–454

Frangakis CF, Rubin DB (2002) Principal stratification in causal inference. Biometrics 58:21–29

Freedman DA, Collier D, Sekhon JS, Stark PB (2010) Statistical models and causal inference. A dialogue with the social sciences. Cambridge University Press, New York

García-Callejas D, Araújo M (2016) The effects of model and data complexity on predictions from species distributions models. Ecol Model 326(Apr):4–12

Gärdenfors P (1988) Causation and the dynamics of belief. In: Harper W, Skyrms B (eds) Causation in decision, belief change and statistics, vol II. Kluwer, Dordrecht, pp. 85–104

Garfinkel A (1981) Forms of explanation : rethinking the questions in social theory. Yale University Press, New Haven

Geels FW (2022) Causality and explanation in socio-technical transitions research: mobilising epistemological insights from the wider social sciences. Res Policy 51(6):104537

Gelcich S, Hughes TP, Olsson P, Folke C, Defeo O, Fernández M, Foale S, Gunderson LH, Rodríguez-Sickert C, Scheffer M, Steneck RS, Castilla JC (2010) Navigating transformations in governance of Chilean marine coastal resources. Proc Natl Acad Sci 107(39):16794–16799

Goodman N (1946) The problem of counterfactual conditionals. J Philos 44:113–128

Greenland S, Pearl J, Robins JM (1999) Causal diagrams for epidemiologic research. Epidemiology 1(10):37–48

Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, Railsback SF, Thulke H-H, Weiner J, Wiegand T, DeAngelis DL (2005) Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310(5750):987–991

Grotzer T (2012) Learning causality in a complex world: understandings of consequence. Rowman and Littlefield Education, Lanham

Guba EG, Lincoln YS (1989) Fourth generation evaluation. Sage Publications, Newbury Park

Hájek A (2007) The reference class problem is your problem too. Synthese 156(3):563–585

Hammersley M (2012) Qualitative causal analysis: grounded theorizing and the qualitative survey. In: Cooper B, Glaesser J, Goom R, Hammersley M (eds) Challenging the qualitative-quantitative divide. Explorations in case-focused causal analysis. Continuum, London, pp. 72–95

Hausman DM, Woodward J (2004) Modularity and the causal Markov condition: a restatement. Br J Philos Sci 55(axh8):147–161

Hazard L, Cerf M, Lamine C, Magda D, Steyaert P (2019) A tool for reflecting on research stances to support sustainability transitions. Nat Sustain 3(2):89–95

Hedström P, Ylikoski P (2010) Causal mechanisms in the social sciences. Ann Rev Sociol 36(1):49–67

Hempel C (1965) Aspects of scientific explanation. The Free Press, New York

Henrich J (2020) The weirdest people in the world : how the west became psychologically peculiar and particularly prosperous. Farrar, Straus and Giroux, New York

Hernan MA, Robins JM (2020) Causal inference: what if. Chapman and Hall/CRC, Boca Raton

Hertz T, Banitz T, Martinez-Peña R, Radosavljevic S, Lindkvist E, Johansson L-G, Ylikoski P, Schlüter M (2024) Eliciting the plurality of causal reasoning in social-ecological systems research. Ecol Soc 29(1)(14)

Hesslow G (1983) Explaining differences and weighting causes. Theoria 49(2):87–111

Hesslow G (1984) What is a genetic disease? On the relative importance of causes. In: Nordenfelt L, Lindahl BIB (eds) Health, disease and causal explanations in medicine. Reidel, Dordrecht, pp. 183–193

Hill AB (1965) The environment and disease: association or causation? J R Soc Med 108(1):32–37

Holland PW (1986) Statistics and causal inference. J Am Stat Assoc 81:945–960

Homer-Dixon T, Walker B, Biggs R, Crépin A-S, Folke C, Lambin EF, Peterson GD, Rockström J, Scheffer M, Steffen W, Troell M (2015) Synchronous failure: the emerging causal architecture of global crisis. Ecol Soc 20(3)

Hruska T, et al (2017) Rangelands as social–ecological systems. In: Briske D (ed) Rangeland systems. Processes, management and challenges, Chap. 8. Springer, Berlin

Hume D (1986/1739) Treatise on human understanding. Penguin, London

Illari P, Russo F (2014) Causality: philosophical theory meets scientific practice. Oxford University Press, Oxford

Illari P, Russo F, Williamson J (2011) Causality in the sciences . Oxford University Press, Oxford

Imbens GW, Rubin DB (2015) Causal inference for statistics, social, and biomedical sciences : an introduction. Cambridge University Press, Cambridge

Johansson L-G (2007) Interpreting quantum mechanics. A realist view in Schrödinger's vein. Ashgate, Aldershot

Johansson L-G (2019) An empiricist view on laws, quantities and physical necessity. Theoria 85(2):69–101

Johansson L-G (2021) Empiricism and philosophy of physics. Synthese Library, no. 434. Springer Nature, Cham

Kant I, Timmermann J, Klemme HF (2003) Die Drei Kritiken. Kritik der Reinen Vernunft. Meiner, Hamburg

Kates RW (2011) What kind of a science is sustainability science? Proc Natl Acad Sci 108(49):19449–19450

Kimmel K, Dee LE, Avolio ML, Ferraro PJ (2021) Causal assumptions and causal inference in ecological experiments. Trends Ecol Evol 36(12):1141–1152

Kolmogorov A (1998/1963) On tables of random numbers. Theor Comput Sci 207(2):387–395

Lade SJ, Niiranen S (2017) Generalized modeling of empirical social-ecological systems. Nat Res Model 30(3)

Lade SJ, Niiranen S, Hentati-Sundberg J, Blenckner T, Boonstra WJ, Orach K, Quaas MF, Österblom H, Schlüter M (2015) An empirical model of the Baltic sea reveals the importance of social dynamics for ecological regime shifts. Proc Natl Acad Sci 112(35):11120–11125

Lang DJ, Wiek A, Bergmann M, Stauffacher M, Martens P, Moll P, Swilling M, Thomas CJ (2012) Transdisciplinary research in sustainability science: practice, principles, and challenges. Sustain Sci 7(S1):25–43

Levin S, Xepapadeas T, Crépin A-S, Norberg J, de Zeeuw A, Folke C, Hughes T, Arrow K, Barrett S, Daily G, Ehrlich P, Kautsky N, Mäler K, Polasky S, Troell M, Vincent JR, Walker B (2012) Social-ecological systems as complex adaptive systems: modeling and policy implications. Environ Dev Econ 18(2):111–132

Lewis D (1973) Counterfactuals. Blackwell, Oxford

Lindegren M, Mollmann C, Nielsen A, Stenseth NC (2009) Preventing the collapse of the Baltic Cod stock through an ecosystem-based management approach. Proc Natl Acad Sci 106(34):14722–14727

Lindkvist E, Basurto X, Schlüter M (2017) Micro-level explanations for emergent patterns of self-governance arrangements in small-scale fisheries—a modeling approach. PLoS One 12(4):e0175532

Lipton P (1991) Inference to the best explanation. Routledge, London

Loehle C (1990) A guide to increased creativity in research: inspiration or perspiration? BioScience 40(2):123–129

Machamer P, Darden L, Craver C (2000) Thinking about mechanisms. Philos Sci 57:1–25

Mackie J (1965) Causes and conditions. Am Philos Q 2(4):245–264

Mackie J (1974) The cement of the universe: a study of causation. Clarendon Press, Oxford

Martinez-Pena R (2023) Analysis of causal argumentation in social-ecological systems research

Martinez-Pena R, Ylikoski P (2023) Constructing the Coleman Boat – mechanism-based theorising in socio-ecological research

Mason J (2018) Qualitative researching, 3rd edn. Sage Publications, Los Angeles

Maxwell JA (2004) Using qualitative methods for causal explanation. Field Methods 16(3):243–264

Maxwell JA (2012) The importance of qualitative research for causal explanation in education. Qual Inq 18(8):655–661

Maxwell JA (2021) The importance of qualitative research for investigating causation. Qual Psychol 8(3):378–388

Menzies P (2012) The causal structure of mechanisms. Stud Hist Philos Mod Sci C 43(4):796–805

Menzies P, Price H (1993) Causation as a secondary quality. Br J Philos Sci 44(2):187–203

Menzies P, Beebee H (2020) Counterfactual theories of causation. In: Zalta EN (ed) The Stanford encyclopedia of philosophy, winter 2020 edn. Metaphysics Research Lab, Stanford University

Mesjasz C (2010) Complexity of social systems. Acta Phys Pol A 117(4):706–715

Miles MB, Huberman AM (1994) Qualitative data analysis: an expanded sourcebook. Sage Publications, Thousand Oaks

Newton I, Cohen IB, Whitman AM (1687/1999) The principia: mathematical principles of natural philosophy. University of California Press, Berkeley

Neyman J (1923) On the application of probability theory to agricultural experiments. Essay on principles. Section 9. Roszniki Nauk Rolniczych 10:1–51

Norström AV, Cvitanovic C, Löf MF, West S, Wyborn C, Balvanera P, Bednarek AT, Bennett EM, Biggs R, de Bremond A, Campbell BM, Canadell JG, Carpenter SR, Folke C, Fulton EA, Gaffney O, Gelcich S, Jouffray J-B, Leach M, Tissier ML, Martín-López B, Louder E, Loutre M-F, Meadow AM, Nagendra H, Payne D, Peterson GD, Reyers B, Scholes R, Speranza CI, Spierenburg M, Stafford-Smith M, Tengö M, van der Hel S, van Putten I, Österblom H (2020) Principles for knowledge co-production in sustainability research. Nat Sustain 3(3):182–190

Norton SB, Cormier SM, Suter II, GW (2014) Ecological causal assessment. CRC Press, Boca Raton

Ostrom E (2007) A diagnostic approach for going beyond panaceas. Proc Natl Acad Sci 104(39):15181–15187

Pearl J (2000) Causality. Models, reasoning, inference. Cambridge University Press, Cambridge

Pearl J (2009) Causal inference in statistics: an overview. Stat Surv 3(Jan): 96–146

Phillips AW (1958) The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. Economica, New Series 25(100):283–299

Preiser R, Schlüter M, Biggs R, García MM, Haider J, Hertz T, Klein L (2021) Complexity-based social-ecological systems research: philosophical foundations and practical implications. The Routledge handbook of research methods for social-ecological systems. Routledge, pp. 27–46

Radosavljevic S, Banitz T, Grimm V, Johansson L-G, Lindkvist E, Schlüter M, Ylikoski P (2023) Dynamical systems modeling for structural understanding of social-ecological systems: a primer. Ecol Complexity 56(Dec)

Railton P (1981) Probability, explanation, and information. Synthese 48(2):233–256

Reichenbach H, Reichenbach M (1999) The direction of time. Dover, Mineola

Ross D, Spurrett D (2007) Notions of cause: Russell's thesis revisited. Br J Philos Sci 58(1):45–76

Rubin DR (2004) Direct and indirect causal effects via potential outcomes (with discussion). Scand J Stat 31:161–179, 195–198

Rubin DB (2005) Causal inference using potential outcomes: design, modelling, decisions. J Am Stat Assoc 100(469):322–331

Russell B (1913) On the notion of cause. Proc Aristotelian Soc 13:1–26

Salmon WC (1984) Scientific explanation and the causal structure of the world. Princeton University Press, Princeton

Salmon W (1997) Causality and explanation: a reply to two critiques. Philos Sci 64(3):461–477

Salmon W (2001) A realist account of causation. In: Marsonet M (ed) The problem of realism. Ashgate, Burlington, pp. 106–133

Samuelson P (1983) Foundations of economic analysis, enlarged edn. Harvard University Press, Cambridge

Schlüter M, Hertz T, García MM, Banitz T, Grimm V, Johansson L-G, Lindkvist E, Rodrigo M-P, Radosavljevic S, Wennberg K, Ylikoski P (2023a, November) Navigating causal reasoning in sustainability science

Schlüter M, Brelsford C, Ferraro PJ, Orach K, Qiu M, Smith MD (2023b) Unraveling complex causal processes that affect sustainability requires more integration between empirical and modeling approaches. Proc Natl Acad Sci 120(41)

Schünemann H, Hill S, Guyatt G, et al (2010) The GRADE approach and Bradford Hill's criteria for causation. J Epidemiol Commun Health 65(5):392–395

Skyrms B (1980) Causal necessity. Yale University Press, New Haven

Solomonoff G (1964a) A formal theory of inductive inference part 1. Inf Control 7(1):1–22

Solomonoff G (1964b) A formal theory of Inductive Inference Part 2. Inf Control 7(2):224–254

Stuart AJ (2014) Late quaternary Megafaunal extinctions on the continents: a short review. Geol J 50(3):338–363

Van Fraassen BC (1980) The scientific image. Oxford University Press, Oxford

van Kerkhoff L, Munera C, Dudley N, Guevara O, Wyborn C, Figueroa C, Dunlop M, Hoyos MA, Becerra L, Castiblanco J (2019) Towards future-oriented conservation: managing protected areas in an era of climate change. Ambio 48(7):699–713

Waernbaum I (2008) Covariate selection and propensity score specification in causal inference. Umeå Universitet, Umeå

West S, van Kerkhoff L, Wagenaar H (2019) Beyond linking knowledge and action: towards a practice-based approach to transdisciplinary sustainability interventions. Policy Stud 40(5):534–555

Witte J, Didelez V (2019) Covariate selection strategies for causal inference: classification and comparison. Biom J 61:1270–1289

Wittgenstein L, Wright GH von, Anscombe GEM (1969) Über Gewissheit: on certainty. Blackwell, Oxford

Woodward J (1997) Explanation, invariance, and intervention. Philos Sci 64(Supplement. Proceedings of the 1996 Biennal Meeting of the Philosophy of Science Association. Part II. Symposia Papers):S26–S41

Woodward J (2002) What is a mechanism? A counterfactual account. Philos Sci 69:S366–S377

Woodward J (2003) Making things happen: a theory of causal explanation. Oxford University Press, Oxford

Woodward J (2008) Invariance, modularity, and all that: cartwright on causation. In: Hartmann S, Hoefer C, Bovens L (eds) Nancy Cartwright's philosophy of science. Routledge, London

Woodward J (2016) Causation and manipulability. In: Zalta EN (ed) The Stanford encyclopedia of philosophy, winter 2016 edn. Metaphysics Research Lab, Stanford University

Ylikoski P (2013) Causal and constitutive explanation compared. Erkenntnis 78:277–297

Index