UNIVERSITY OF CALIFORNIA SAN DIEGO

The Philosophical Significance of Causal Diversity

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Philosophy

by

Weixin Cai

Committee in charge:

Professor Nancy Cartwright, Chair
Professor David Danks
Professor Stephan Haggard
Professor Karen Kovaka
Professor Kerry McKenzie

2026

The Dissertation of Weixin Cai is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2026

DEDICATION

To my parents, Lihua Du and Quidong Cai.

I would not have gone this far without their deepest love and trust.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

When I am silent, I feel replenished; as I open my mouth, I sense emptiness.

Lu Xun, "Inscription," *Wild Grass*

As I sit at my desk, thinking about what words to leave here, I find it difficult to decide where to begin. The past sixteen years, from the time I started my undergraduate studies to the present, feel like a long yet fleeting dream. What made the dream possible were my parents, who did not receive a college education and knew little about the range of academic majors, nevertheless trusted me wholeheartedly. They encouraged me to choose the field I wished to study and to pursue the academic path I wanted to follow, even though this meant that I would have to move farther and farther away from home—from Quanzhou in southeastern China to Beijing, then to Canada, and finally to the United States—with less and less time to spend with them. Each time I returned home during the summer break, I noticed that my parents had grown older, both physically and cognitively. The thought of this always fills me with both gratitude and sorrow. Now, at last, it is time to bring this long journey to an end and return home to the faces that have always welcomed me.

During my academic journey, I am fortunate enough to get to know and learn from many people. I would like to first thank both of my advisors, Nancy Cartwright and David Danks. When I was an undergraduate student in China, only a relatively small number of works in philosophy of science had

been translated into Chinese and were accessible to ordinary philosophy students. Nancy's works were among them. She was one of the most well-known philosophers whom Chinese students interested in philosophy of science would have heard of. At that time, I would never have imagined that one day I would pursue my doctoral studies under her supervision. Life is full of surprises. Later, during my master's studies in Canada, I became increasingly interested in causation. This interest led me to UC San Diego. I still remember my excitement when she agreed to be my co-advisor. Many themes in this dissertation, including middle-range theories, causal diversity, background conditions, and Simpson's paradox, have been significantly influenced by my reading of her work.

David was one of the first faculty members I met in person after completing my first year of doctoral study online and finally arriving in San Diego in 2021. Soon after we met, he immediately agreed to my request for regular meetings, despite his busy schedule. Whenever I shared my thoughts with him, I was always impressed by his ability to guide me toward new and inspiring ideas I had not considered, and to articulate complicated thoughts in a highly accessible way. This admiration eventually led me to the somewhat ambitious idea of inviting him to become one of my co-advisors. Over the years, he and Nancy have set a high standard for me to follow, not only in scholarship but also in professionalism. As an old Chinese saying goes, "The lofty mountain stands before me to look up to; the righteous road lies before me to follow. Although I may not reach the destination, my heart will always yearn for it." (高山仰止，景行行止。虽不能至，然心向往之。) I am deeply indebted to them.

I would also like to thank the other members of my committee, Stephan Haggard, Karen Kovaka, and Kerry McKenzie, both for their willingness to serve on my committee and for the support I have

received from them in other ways. I am very grateful to Steph for his encouragement and kind words about my final paper when I took his graduate seminar on qualitative causal inference, to Karen for accepting my invitation to join my committee without hesitation under such unusual circumstances, and to Kerry for her consistent and warm support over the years as the director of graduate studies, as well as for the opportunities she offered me, including inviting me to serve as a commentator at Pacific APA 2021 and to work as her summer research assistant.

I am deeply grateful to both the department and the city. Over the past six years, I have been consistently supported by the department in many respects. The faculty members and administrative staff have been among the most understanding people I have ever known. Whenever I sought help, they always responded to my questions promptly and kindly. I have benefited a great deal from my time here, both academically and personally. Although this may sound like a cliché, I remain deeply thankful to the department for giving me the opportunity to begin my journey here six years ago.

San Diego is one of the most comfortable cities I have ever lived in. Part of this comfort comes from its familiarity. The sea, sunshine, and breeze in San Diego frequently remind me of my hometown, Quanzhou, a coastal city in southeastern China on the other side of the North Pacific Ocean. Another part of this comfort comes from its uniqueness. San Diego has welcomed me with its vibrant cultural life, distinctive natural scenery, Californian urban aesthetics, and lovely though sometimes noisy seals and sea lions. I have begun to miss my time here even before I leave.

Now my long journey in San Diego is coming to an end. I will cherish the lovely memories of everyone I have known here, with the sincere hope that we will reconnect in the near future.

As the closing of one chapter of my life, the completion of my doctoral study is itself an instance of joint causation, with the help of many people making a joint causal contribution to this outcome. It is beyond my ability to list everyone whom I should thank. But please allow me to single out my previous advisors and mentors, Holly Andersen, Eddy Keming Chen, Donghui Han, Chang Liu, Xiaoli Liu, and Jennifer Wang. I am indebted to them for their support at different stages of my academic growth.

I would also like to thank my friends in the reading group we formed in 2019. Over the past seven years, we have read papers together, commented on each other's work, explored potential opportunities for collaboration, shared updates about our lives, and offered one another support and sympathy in weekly or biweekly meetings. Although many members of the group have entered, or are about to enter, a new stage of life, I hope our friendship will continue.

The past two months have been the most chaotic period of my life. Some unexpected hurdles arose in my job search and significantly disrupted my career plan. I did not share my feelings with others, including my parents, because I did not want them to worry about me. My girlfriend, Yayun Yang, is the only person who understands my circumstances and emotions. She knows well my anxiety, dissatisfaction, loss of confidence, and many other difficult feelings that I have experienced during this period. Although she is also on the job market this year and faces no less uncertainty than I do, she always

responds to my feelings with patience and encourages me to remain confident in myself. I am deeply grateful for all her love and support.

Lastly, materials from the following chapters have been published or submitted for publication. I am grateful to my co-authors for kindly allowing me to incorporate them into my dissertation.

Chapter 1, in full, is a reprint of the material currently under review at *Philosophy Compass*.

Chapter 3, in full, is a reprint of the material co-authored with Zili Dong, which is accepted for presentation at PSA 2026 and conditionally accepted for publication in *Philosophy of Science*. The dissertation author was the corresponding and co-first author of this paper.

Chapter 4, in full, is a reprint of the material co-authored with Zili Dong and Shimin Zhao, as it appears in *European Journal for Philosophy of Science*, 2024. The dissertation author was the corresponding and co-first author of this paper.

VITA

2014 Bachelor of Arts in Philosophy, Renmin University of China

2017 Master of Arts in Philosophy, Renmin University of China

2019 Master of Arts in Philosophy, Simon Fraser University

2026 Doctor of Philosophy in Philosophy, University of California San Diego

PUBLICATIONS

In English

1. Cai, Weixin and Zili Dong. "Causal-Structural Explanation." Conditionally accepted by *Philosophy of Science*.

2. Dong, Zili, Weixin Cai and Shimin Zhao (2024). "Simpson's Paradox Beyond Confounding." *European Journal for Philosophy of Science* 14, 44: 1–22.

3. Cai, Weixin (2018). "Causal Exclusion and Causal Autonomism." *Frontiers of Philosophy in China* 13 (3): 402–419.

In Chinese

4. Cai, Weixin (forthcoming). "Simplicity: A Secondary Ultimate Cognitive Value for Non-Ideal Cognitive Agents" (简单性：非理想认知主体的次要终极认知价值).*Journal of Dialectics of Nature* (自然辩证法通讯).

5. Cai, Weixin (2025). "Determinism as a Stance" (作为立场的决定论). *Journal of Dialectics of Nature* (自然辩证法通讯) 47 (10): 48–54. Excerpt reprinted in *Chinese Social Science Digest* (中国社会科学文摘) 2025 (12): 152.

ABSTRACT OF THE DISSERTATION

The Philosophical Significance of Causal Diversity

by

Weixin Cai

Doctor of Philosophy in Philosophy

University of California San Diego, 2026

Professor Nancy Cartwright, Chair

"Causal diversity," as an umbrella term, refers to the diversity found within causation and related subject matters. This dissertation investigates four specific topics on which considerations about causal

diversity may shed new light: philosophical approaches to theorizing about causation, background conditions, causal explanations, and the causal grounds of Simpson's paradox.

First, the dissertation distinguishes between two kinds of philosophical accounts of causation: grand theories and middle-range theories. Grand theories of causation treat causation primarily as a homogeneous category and aim to improve our philosophical understanding of causation at a highly general level. By contrast, middle-range theories of causation treat causation primarily as a heterogeneous category, with each theory aiming to improve our philosophical understanding about a specific type of causal relation. The two kinds of theory complement each other in providing a comprehensive philosophical understanding of causation.

Second, the dissertation investigates the extent to which the background conditions under which causes produce their effects are diverse. By offering a focused analysis on facilitators, i.e., background conditions that help a cause produce its effect more efficiently without affecting whether it causes that effect, the dissertation argues that many apparently diverse types of background condition can be reduced to enablers. This suggests that background conditions are not as diverse as they initially appear to be.

Third, the dissertation offers an account of causal-structural explanation, a specific kind of scientific explanation that explains correlations or regularities among causal factors in terms of the underlying causal structure that gives rise to them. Causal-structural explanations challenge the received view of causal explanation because this view does not specify whether an explanation is causal in virtue of its explanans or in virtue of the dependence relation between the explanans and the explanandum.

Fourth, the dissertation argues against the view that Simpson's paradox in causal inference can be fully addressed by identifying confounders in the relevant graphical causal models. There is no unified solution to Simpson's paradox because the paradox may arise from causal grounds that do not involve confounding.

Introduction

Causation is often recognized as a homogeneous category whose members share core ontological or epistemic features. This recognition has underwritten the longstanding philosophical effort to identify those common features. For instance, many philosophers assume that causation has a single "nature" or "essence," and take one central aim of the metaphysics of causation to be the identification of the ontological features that constitute the nature of causation. Efforts in the pursuit of this aim have yielded a variety of metaphysical theories of causation, including regularity theories, counterfactual theories, process theories, manipulability theories, and many others. Similarly, a lot of methodologically oriented philosophers regard interventionism as a unified analysis of causation that can illuminate valid methods of causal inference across different contexts.

The widespread philosophical tendency to treat causation as a unified subject matter seems at odds with the profound causal diversity found in science. For one thing, each scientific discipline is primarily concerned with causal relations and causal structures involving its distinctive set of objects, properties, events, and processes. The specific features of the entities studied in each discipline shape not only the characteristics of relevant causal relations but also the methods appropriate for inferring them. For another, even within a single discipline, causal relations often vary considerably along dimensions that have received relatively little attention in traditional philosophical discussions of causation, such as rate of change and whether a cause exerts a linear or monotonic influence on its effect. What philosophical lessons should we draw from causal diversity displayed in science?

In the dissertation, I explore the philosophical consequences of picturing causation as a diverse category. In particular, I discuss the implications of causal diversity on how we may develop philosophical theories of causation (Chapter 1) and demonstrate middle-range causal theorizing, the theorizing approach capable of addressing causal diversity, by offering an in-depth analysis for the kind of causal factors that I call facilitators (Chapter 2). In addition, I offer two case studies to illustrate how considerations related to causal diversity can shed light on relevant philosophical discussions. One concerns diversity in causal explanation. It challenges the traditional view of causal explanation by arguing that this view cannot accommodate a distinctive kind of explanation that I call "causal-structural explanation" (Chapter 3). The other casts doubt on the claim that Simpson's paradox in causal inference admits of a unified solution by showing that the paradox can arise from diverse causal grounds (Chapter 4).

More specifically, Chapter 1 provides a survey of two general approaches to theorizing about causation in philosophy. Grand theories of causation treat causation primarily as a unified category and aim to identify important shared features of causal relationships. In Part I of this chapter, I offer a general characterization of the most prominent variant of grand theory of causation—objectivistic grand metaphysical theories of causation—and survey three main challenges these theories face in the literature, advanced respectively by (a) causal particularism, (b) causal instrumentalism, and (c) causal functionalism. In addition, I briefly discuss attempts to develop grand epistemological theories of causation and the main challenge they confront.

Causation is diverse in many ways. Grand theories of causation are ill-suited to illuminate the distinctive characteristics of scientifically significant varieties of causation. In Part II of Chapter 1, I identify four major kinds of causal diversity: (a) the diversity of the non-essential features of causal relationships, (b) the diversity of causal systems, (c) the diversity of causal domains, and (d) the diversity of causal roles. I then articulate the notion of middle-range theory of causation, which is better suited than grand theories to addressing kinds of causal diversity found in science, and distinguish its three major variants: feature-centered, domain-centered, and role-centered theories of causation.

Most philosophical discussions of background conditions have focused on enablers, which make it possible for a cause to produce its effect in actual situations. In Chapter 2, I offer a role-centered, middle-range account of causation by examining an apparently distinct class of background conditions called facilitators. Facilitators help a cause produce a target effect *more efficiently*, without making an essential difference to *whether* the cause produces the effect. I argue for two theses. First, despite their seemingly distinctive causal role, facilitators are reducible to enablers that redundantly enable a cause to produce its effect when other existing enablers would already suffice to do this job. Second, mainstream difference-making accounts of causation fail to identify facilitators as causes of the facilitated effects. These theses have implications both for our understanding of background conditions in general and for the causal status of enablers in particular. On one hand, besides facilitators, many seemingly distinct kinds of background conditions are also reducible to redundant enablers. This justifies a partially unified picture of background conditions. On the other hand, if facilitators are redundant enablers without being causes, they challenge the claim that all enablers are causes in the proper sense. To defend this claim

from a difference-making perspective, causes should be understood as difference-makers for the *actual causal processes* leading to their effects.

Chapter 3 concerns primarily with the diversity of causal explanations instead of causal relations. The received view of causal explanation holds that to causally explain a phenomenon is to provide information about its relevant causes. However, this view is ambiguous as to whether an explanation is causal primarily because of (a) the explanans it cites or (b) the dependence relation it posits between the explanatory relata. These options diverge on whether causal-structural explanations should count as causal explanations. To show this, my co-author and I analyze the explanatory pattern of such explanations and argue that they have a dual nature: while the explanans provides causal information to account for the explanandum, the explanandum nevertheless non-causally depends on the explanans.

Chapter 4 shifts its focus to causal inference and explores how it can bear on considerations about causal diversity. Simpson's paradox (SP) is a statistical phenomenon where the association between two variables reverses, disappears, or emerges, after conditioning on a third variable. It has been proposed (by, e.g., Judea Pearl) that SP should be analyzed using the framework of graphical causal models (i.e., causal DAGs) in which SP is diagnosed as a symptom of confounding bias. This chapter contends that this confounding-based analysis cannot fully capture SP: there are cases of SP that cannot be explained away in terms of confounding. Previous works have argued that some cases of SP do not require causal analysis at all. Despite being a logically valid counterexample, my co-authors and I argue that this type of cases poses only a limited challenge to Pearl's analysis of SP. In our view, a more powerful challenge to Pearl comes from cases of SP that do require causal analysis but can arise without confounding. We

demonstrate with examples that accidental associations due to genetic drift, the use of inappropriate aggregate variables as causes, and interactions between units (i.e., inter-unit causation) can all give rise to SP of this type. The discussion is also extended to the amalgamation paradox (of which SP is a special form) which can occur due to the use of non-collapsible association measures, in the absence of confounding.

Chapter 1

Two Approaches to Theorizing about Causation

Part I: Grand Theories of Causation and Their Discontents

1. Introduction

Compare the two causal relationships:

(1) HIV infection causes AIDS.

(2) An increase in tariffs on imported electric vehicles causes a decline in their sales.

The two relationships presumably share certain important features by virtue of which both are *causal* in nature, even though philosophers disagree on what these features are. For example, proponents of the regularity theory of causation believe that both causal relationships exhibit a stable, non-accidental kind of regularity between the occurrence of the cause and that of the effect, while advocates of the counterfactual theory claim that both involve a counterfactual dependence of the effect on the cause.

Nevertheless, these causal relationships differ significantly in some other important respects. First, HIV infection is a *necessary* causal condition for AIDS, because no individual can develop AIDS without first being infected with HIV. By contrast, raising tariffs on imported electric vehicles is not a

necessary causal condition for a decline in their sales. Even if tariffs remain unchanged, sales may still decrease due to other factors, such as a growing consumer preference for domestically produced vehicles.

Second, the causal relationship between HIV infection and AIDS is highly *deterministic*: an individual infected with HIV is almost certain to develop AIDS, in the absence of timely and effective treatment such as antiretroviral therapy. By contrast, the causal relationship between tariff increases and a decline in imported electric vehicle sales is considerably more probabilistic, since its underlying mechanism is susceptible to interference from many other factors.

Third, the two causal relationships pertain to different *domains of reality*. The causal relationship between HIV infection and AIDS is an instance of biological causation, whereas the one between tariff increases and a decline in imported electric vehicle sales is an instance of economic causation.

In a similar way, the full range of causal relationships exhibits both unity and diversity. Like atoms, birds, and paintings, all causal relationships seem to share certain salient features which bind them into a relatively *unified*—though not necessarily monistic—category. Yet, as with the existence of different types of atom, bird, and painting, there are also significant differences among causal relationships that warrant their further classification into *distinct* types or sub-categories.

Causation plays a key role both in science and in philosophy. For scientists, acquiring reliable causal knowledge is essential for achieving central scientific goals such as making accurate predictions (Broadbent, 2015), providing plausible explanations (Ross & Woodward, 2023), and designing effective

interventions (Cartwright & Hardie, 2012). For philosophers, the notion of causation is not only intriguing by itself but also deeply intertwined with a wide range of philosophical topics—confirmation (Wheeler & Scheines, 2013; Zhao, 2024), knowledge (Goldman, 1967; Dretske, 1981), mental content (Stampe, 1977; Rupert, 2008; Adams & Aizawa, 2021), moral responsibility (Hart & Honoré, 1985; Sartorio, 2013), perception (Grice, 1961; Ganson, 2021), physicalism (Kim, 2005; List & Menzies, 2009), scientific explanation (Machamer et al., 2000; Woodward, 2021a; Ross, 2025), to name only a few—such that a good understanding of causation can significantly inform debates over these topics. Consequently, philosophers, especially metaphysicians and philosophers of science, are strongly motivated to theorize about causation, aiming to develop successful theories that can enhance our philosophical understanding of causation.

The kind of philosophical understanding a theory of causation can provide is significantly shaped by the range of causal relationships it takes to be its theoretical scope, and by the questions it aims to answer about them. Given that causation exhibits both unity and diversity, two approaches to theorizing about causation emerge naturally. One approach brings the full range of causal relationships into the theoretical scope and investigates the relevant conceptual, metaphysical, epistemological, or pragmatic questions that concern causation at a highly general level: for instance, what the concept "cause" really means, what essential features causal relationships have, and on what evidential grounds we can infer that one factor causes another. In pursuing these questions, this approach treats causation primarily as a unified, or homogenous, category in which all (or at least a significant majority of) causal relationships share certain important features. The main goal of this approach is to identify what these

shared features are, or to examine whether a given feature is in fact shared by all, or at least most, causal relationships. I will refer to this unity-based approach to theorizing about causation as "grand causal theorizing" and to its products as "grand theories of causation."

By contrast, the other approach to causal theorizing treats causation primarily as a diverse, or heterogenous, category in which causal relationships vary in certain important ways, e.g., in whether they exhibit features such as causal necessity and causal determinacy, and in the domains of reality to which they belong. The goal of this approach is thus to develop theories that address philosophical questions specific to causal relationships with certain distinctive features, rather than questions uniformly applied to all causal relationships. For instance, one might develop a theory tailored to the class of causal relationships in which the cause functions as a necessary condition for its effect, by examining the conceptual, metaphysical, epistemological, or pragmatic issues that specifically arise for this kind of causal relationship. Because theories of causation developed under such a diversity-based approach will adopt a much narrower theoretical scope than grand theories of causation, I will refer to this approach as "middle-range causal theorizing" and to its products as "middle-range theories of causation."

Grand causal theorizing prevails in current philosophical literature on causation. Most of Part I will focus on its most prominent variant, *objectivistic grand metaphysical theorizing* about causation, including a general characterization of this kind of causal theorizing (Section 2) as well as some main challenges it faces (Section 3). Besides, I will also cover attempts to build grand epistemological theories

of causation and its main challenge (Section 4). In Part II of this chapter, I will survey several ways in which causation is diverse, which justify the need for more middle-range theories of causation.

2. Objectivistic Grand Metaphysical Theorizing about Causation

Most philosophical efforts to develop grand theories of causation have been devoted to uncovering the *nature*, or *essence*, of causation, namely, the set of ontological features that causal relationships essentially possess. Such theories typically address highly general metaphysical questions about causation, including:

- What features are constitutive of causal relationships and distinguish them from non-causal ones? (Lewis, 1973; 2000; Mackie, 1974; Salmon 1984; Dowe 2000; Woodward, 2003; Mumford & Anjum, 2011)?
- What kinds of entity can serve as causal relata (Kim, 1973; Lewis, 1986a; Menzies, 1989; Mellor, 1995; Paul, 2000; Woodward, 2003)?
- Must causal relationships be temporally asymmetric (Dummett, 1964; Price, 1995), or transitive (Lowe, 1980; Hall, 2000; Hitchcock, 2001; McDonnell, 2018)?
- How are token-level and type-level causal relationships ontologically connected? (Davidson, 1967; Sober, 1984; Cartwright, 1989; Elles, 1991; Hitchcock, 1995)?

These questions are formulated to range over all causal relationships. In addition, it is often assumed that the full range of causal relationships are sufficiently unified with respect to their core metaphysical

characteristics that the correct answers to these questions are uniform enough to be captured by grand, systematic theories of causation.

Furthermore, most philosophers presuppose an *objectivistic* view of causation (I will call "causal objectivism" hereafter) when engaging with these questions. On this view, causation is an objective feature of reality in the sense that causal relationships are not projected onto the world by epistemic agents such as humans, but obtain independently of any individual's or group's subjective states (causal beliefs, conceptual frameworks, manipulative agency, etc.). This commitment to the ontological objectivity of causal relationships has three benefits for the development of grand metaphysical theories of causation: first, it underwrites the objectivity of scientific endeavor to discover causal relationships of interest, thus offering naturalistically inclined philosophers a well-grounded motivation for theorizing about causation; second, it commits theorists to an objective target domain of causal relationships, thereby precluding the relativist view that competing grand theories of causation could be equally true simply because they concern different, subjectively defined ranges of causal relationships; third, it suggests that there is an objective metaphysical basis for distinguishing causal from non-causal relationships, therefore making the search for that ground a realistic aim for grand metaphysical theorizing about causation.

For sake of ease, I will refer to the practice of developing a theory of causation that (i) takes the *full* range of causal relationships as its theoretical scope, (ii) aims to address highly general *metaphysical* questions about causation, and (iii) assumes *causal objectivism* as "objectivistic grand metaphysical theorizing about causation."

Most well-known metaphysical theories of causation exemplify objectivistic grand metaphysical theorizing, including various forms of probabilistic theories (Reichenbach, 1956; Suppes, 1970; Cartwright, 1979; Eells, 1991; Glynn, 2011), counterfactual theories (Lewis, 1973; 2000; Yablo, 2002; Sartorio, 2005), process theories (Fair, 1979; Salmon, 1984; Dowe, 2000), mechanistic theories (Glennan, 1996), dispositional theories (Ellis, 2001; Molnar, 2003; Mumford & Anjum, 2011; Bird, 2020), as well as hybrid theories that analyze causation in terms of a combination of components drawn from the above approaches (Paul, 2000; Schaffer, 2001; Loew, 2019).[1] While these theories disagree about which features are essential to causation, they share a global theoretical scope on causation and treat it primarily as a sufficiently unified category with respect to its essential metaphysical features. They also share three further characteristics typical of grand theories of causation.

First, key theoretical statements in these theories are often *universal generalizations* that apply to the full range of causal relationships. For instance, probabilistic theories claim that causes are factors that raise the probabilities of their effects under suitable conditions. This is a universal generalization intended to range over all causal relationships.

[1] Regularity theories (Hume, 1739; Mill, 1843; Mackie, 1974; Baumgartner, 2013; Baumgartner & Falk, 2023) and agency theories (Menzies & Price, 1993) hold that causal relationships essentially involve human perspectives and interests projected onto certain objective relations in the world, such as robust regularities or relations by which free agents can effectively bring about targeted outcomes. Although these theories deny that causation is a fully objective feature of the world, they nonetheless posit an objective component in its basis. In this regard, they can be viewed as partially exemplifying objectivistic grand metaphysical theorizing.

Second, as a cost of their global theoretical scope (or the universal domain of their theoretical statements), these theories yield only *coarse-grained* information about particular causal relationships. They abstract away from their non-essential features, including features of the cause, the effect, and the specific manner in which the cause brings about the effect, since such features are not widely shared across causal relationships.

Third, granted that successful philosophical theories can enhance philosophical understanding, a successful grand metaphysical theory of causation will enhance our philosophical understanding of causation *in general*, rather than merely of a restricted sub-class of causal relationships.[2] Causal objectivists further maintain that the right kind of understanding conveyed by a grand metaphysical theory of causation should enable us to grasp the objective metaphysical basis in virtue of which causal relationships possess an ontologically objective status.

Note that objectivistic grand metaphysical theories of causation need not always be monistic. A dualistic theory positing two distinct natures of causation, such as counterfactual dependence and production in Hall (2004), also exhibits the characteristics of grand theories of causation. Although such a theory denies that causation has a single, universally shared nature, it aims to cover the full range of causal relationships, treating them as sufficiently unified to be accounted for by a single theoretical framework. Also, its core theoretical statement—the claim that any causal relationship consists either in

[2] I leave it an open question what the succuss conditions for a philosophical theory consist of.

counterfactual dependence or in production—is a universal generalization that ranges over all causal relationships. If such a dualistic theory turns out to be a successful one, it will enhance our philosophical understanding of causation in general just as successful monistic theories do.

Fairly speaking, objectivistic grand metaphysical theorizing has long dominated philosophical discussions of causation. Nevertheless, this approach has its discontent. Recent work suggests that objectivistic grand metaphysical theorizing is under pressure from two directions. These challenges are typically advanced by philosophers of science who attend closely to how causation is actually assessed in scientific practice. These two lines of challenge can be formulated as the following theses:

- **The Inappropriateness Thesis**: Objectivistic grand metaphysical theories of causation cannot provide an *appropriate* philosophical understanding of causation.
- **The Inadequacy Thesis**: Grand theories of causation (including objectivistic grand metaphysical theories) alone cannot provide an *adequate* philosophical understanding of causation.

I will present the challenges grouped under the Inappropriateness Thesis in the next section, and reserve the discussion of the case for the Inadequacy Thesis for Part II of this chapter.

3. Challenges to Objectivistic Grand Metaphysical Theorizing about Causation

The success of objectivistic grand metaphysical theorizing about causation presupposes (i) that all causal relationships are sufficiently unified in their essential or core metaphysical features, (ii) that those features constitute an objective metaphysical basis that underwrite the ontological objectivity of the causal relationships scientists aim to discover, and (iii) that the kind of philosophical understanding provided by this approach is genuinely valuable. However, all three presuppositions are far from uncontroversial, as they are challenged by causal particularists, causal instrumentalists, and causal functionalists respectively. These philosophers seek to expose weakness in objectivistic grand metaphysical theorizing about causation and to offer alternative visions of how we might better theorize about causation. Taken together, they amount to a systematic skepticism against the *unity*, *objectivity*, and *significance* of the kind of philosophical understanding of causation promised by objectivistic grand metaphysical theorizing.

3.1 The Challenge from Causal Particularism

Consider the following claims:

(3) The wind *pushes* the door closed.

(4) The rain *wets* the ground.

(5) The sun *melts* the ice.

Each claim represents a causal relationship and can be reformulated in the form of "X causes Y" such as the follows:

(3′) The wind *causes* the door's closure.

(4′) The rain *causes* the ground's wetting.

(5′) The sun *causes* the ice's melting.

The fact that the causal relationships represented by (3)–(5) can be described by the same form of causal claim that employs the univocal causal concept "cause" may lead one to think that these causal relationships, as well as many others, share the same essential features. The concept "cause" picks out those shared features, thereby allowing different causal relationships to be subsumed under a single schematic form of causal claim.

Instead, causal particularists deny that there are any generally shared essential features singled out by the very concept "cause" (Anscombe, 1971; Cartwright, 2004). They maintain that "there is an untold variety of causal relation[ships]," each with its unique causal essence (Cartwright, 2004: 814). The genuine essence of a causal relationship is content-rich in the sense that it characterizes the distinctive way in which a specific cause affects a specific effect. For instance, the way in which the wind *pushes* the door closed is significantly different from the way in which the sun *melts* the ice, even though both relationships can be roughly characterized as one thing *causing* another. It is deeply misleading to conceive different causal relationships as sharing the same essence merely because they can be referred to by the same causal concept "cause."

What has driven causal particularists to favor particularized causal essences? One factor is skepticism about the existence of any generally shared essential feature of causation, given the sustained failure to develop an (objectivistic) grand metaphysical theory of causation that can successfully identify such a feature. By far, every proposed grand metaphysical theory of causation struggles to account for certain causal relationships, despite seemingly successful in illuminating some others: regularity theories are well known to have difficulty accommodating irregular causal relationships instantiated only once in history and excluding non-causal regularities such as correlations between two regular effects of a common cause; probabilistic theories have a hard time analyzing causes that lower or do not affect the probabilities of their effects (Salmon, 1971; Hesslow, 1976); counterfactual theories are famously challenged by varieties of counterfactually fragile causal structures such as early and late preemption, overdetermination, double prevention, and trumping (Lewis, 2000); process theories appear ill-suited to cases of causation by absence, where the cause and the effect are not connected by a physical process, and to causal relationships in social or economic contexts, where the appeal to physical processes may seem unnecessary (Hoover, 2001); ...[3] One straightforward explanation for these unsuccessful attempts is simply that there are no generally shared essential features to be identified by a grand metaphysical theory of causation in the first place. Call this line of argument *the argument from prevalent counterexamples*.

Another argument often invoked by causal particularists, *the argument from conceptual primitivity*, supports the existence of diverse and particularized causal essences by appealing to the

[3] For counterexamples to more grand theories of causation, see Cartwright (2004).

conceptual primitivity of "thick" causal concepts that capture particular causal relationships. Both G. E. M. Anscombe (1971) and Nancy Cartwright (2004) highlight that, in addition to the general causal concept "cause," there is an abundance of special, or "thick," causal concepts in our causal vocabulary (e.g., "push," "wet," "melt," "scrape," "carry," "eat," "burn," "squash," "hurt," "open," etc.), which we readily grasp to pick out specific activities by whose performance a cause affects its effect.[4] These thick causal concepts are considered by causal particularists to be conceptually primitive for at least two reasons. For one thing, it is impossible to grasp the general causal concept "cause" without already grasping a host of thick causal concepts. For another, one cannot conceptually reduce thick causal concepts to the general causal concept "cause." To see this, compare (3) with the following claim:

(6) The wind *chills* people and *prompts* them to close the door.

While both (3) and (6) attribute a causal relationship between the wind and the door's closure and thus imply (3'), they express essentially distinct causal relationships between the two factors. Recasting both causal claims into (3') thus omits crucial causal information that distinguishes one from the other. Causal particularists take the conceptual primitivity of thick causal concepts to track the metaphysical fundamentality of the particularized causal essences to which those concepts refer.

Thus, in the view of causal particularism, there is no informative general answer to the question of what features make a relationship causal. This, if true, implies that any attempt at developing grand

[4] These causal verbs are also called "causatives" in Reiss (2015).

metaphysical theories of causation is bound to fail, because it is simply a wild-goose chase, trying to hunt something that does not exist in the first place. We routinely label particular causal relationships with the univocal and abstract concept "cause," mainly because doing so enables us to subsume essentially distinct causal relationships under a single conceptual framework and apply general inference methods to establish causal claims about them, rather than because they have a shared essence. For causal particularists, a more reasonable theorizing approach is to pursue *highly local* metaphysical theories of causation, each tailored to elucidate the nature of a thick causal relationship, such as a theory of causal relationships in which the wind pushes the door's closure, a theory of causal relationships in which the rain wets the floor, etc.

3.2 The Challenge from Causal Instrumentalism

Causal instrumentalists challenge causal objectivism widely shared by those engaging in grand metaphysical theorizing about causation (Norton, 2003; 2024; Williamson, 2005; 2006a; 2006b). For causal instrumentalists, causation is not an objective feature of reality, but rather an epistemic category imposed onto reality by epistemic agents like humans merely for its *utility*. By representing events in terms of causes and effects, we can grasp the processes that connect distinct events more conveniently, and devise effective strategies to bring about certain outcomes. But, as causal instrumentalists warn, the epistemic utility of causal representations should not mislead us into thinking that causal relationships exist independently of our causal beliefs or perspectives.

Similar to causal particularists, causal instrumentalists appeal to the argument from prevalent counterexamples to make their case. The difference is that they identify the locus of the persistent failures of grand metaphysical theories of causation in terms of causal objectivism, which such theories typically presuppose. If causation is an integral part of the objective reality, the set of relationships we label "causal" should share some essential features which constitute the objective metaphysical basis that distinguish causal relationships from non-causal ones. Such features, if exist, will set up a consistent factual constraint on what relationships we should label "causal" (Norton, 2003), or prescribe a consistent evidential basis for us to reasonably judge some relationships as causal and others as non-causal (Williamson, 2006a; 2006b), especially in scientific contexts where such practices are regarded as more well-considered.

However, the history of science suggests that scientists routinely label newly discovered relationships that do not conform to pre-existing grand conceptions of causation "causal." For instance, causation was once firmly believed by scientists to be essentially deterministic. This conviction was largely abandoned when developments in quantum mechanism suggested that certain microphysical events are not strictly determined by the antecedent events that scientists identified as their causes. Similarly, scientists also appeal to a highly divergent set of information as evidential indicators for causation, such as "observed associations, observed independencies, temporal cues, known mechanisms, theoretical connections, experiments, controlled trials, other causal knowledge, intuitions about subjunctive conditionals," etc. (Williamson, 2006a: 69).

In light of persistent counterexamples and diverse evidential indicators found in science, causal instrumentalists believe that we can even make a "pessimistic induction": even if grand metaphysical theories of causation continue to develop, new scientific counterexamples to these theories will likely arise, along with perhaps new evidential indicators that they cannot accommodate. These considerations suggest that the notion of causation is in fact incapable of imposing any factual or evidential constraint in science. Instead, it has become "so plastic that virtually any new science can be made to conform to it" (Norton, 2003: 2). The lack of nontrivial factual and evidential content in the notion of causation thus implies that there is no defining feature associated with those relationships we label "causal."

Thus, contrary to those adopting an objectivistic view of causation, causal instrumentalists insist that an adequate grand metaphysical theory of causation should highlight its epistemic and instrumental nature. For instance, Williamson (2005; 2006a; 2006b) offers a grand metaphysical account of "epistemic" causality, which analyses the nature of causation in terms of rational causal beliefs. These causal beliefs are best supported by one's causally relevant evidence and can help one achieve a number of pragmatic goals. By contrast, Norton (2003; 2024) proposes a heuristically useful "folk theory" of causation. This theory resembles a grand objectivistic metaphysical theory in its content, but crucially differs from it in aiming merely to articulate a set of fallible causal features that epistemic agents may adopt to facilitate causal cognition.

3.3 The Challenge from Causal Functionalism

Causal functionalists do not deny the possibility of developing objectivistic grand metaphysical theories of causation. Rather, they question whether the philosophical understanding of causation provided by such theories has much pragmatic significance for us (Glymour et al., 2010; Danks, 2013; Woodward, 2014; 2015a; 2021b). Causal functionalism starts with the idea that causal cognition, or the epistemology of causation, serves important functions in promoting pragmatic goals such as prediction, explanation, and manipulation. From this perspective, it is pragmatically important to identify the cognitive tools or inference strategies we (i.e., humans) actually use in our causal cognition, both in everyday scenarios and in scientific inquiry (the *descriptive* project), and to evaluate how well they promote these goals (the *normative* project) (Woodward, 2021b).

For causal functionalists, it is unclear how prevailing grand metaphysical theories of causation can enhance our functional understanding of causal cognition. Some grand metaphysical theories of causation may have methodological implications about how we may establish a causal claim. For example, if the probabilistic theory of causation, which identifies causation with probability-raising relationship under suitable background conditions, is correct, it implies that evidence about whether C raises the probability of E will of utmost importance in our inference about whether C causes E. However, most grand metaphysical theories of causation, especially those postulating heavy-duty metaphysical entities like possible worlds, powers, or universals, are often disconnected from scientific practices about causal

inference, and have few implications on why scientists have been so successful in searching for real-world causal relationships or how scientists should view causation when they are conducting actual research.

By contrast, causal functionalists recommend developing "functional" accounts of causation that illuminate how we engage in causal cognition in both ordinary and scientific scenarios and help us evaluate how well different forms of causal cognition serve our pragmatic goals (Woodward, 2014). A representative example is interventionism. By treating successful intervention in the world as a central pragmatic goal, this account understands causal relationships as relationships exploitable for intervention purposes (Woodward, 2003; 2021b). It offers a basis for understanding how not only ordinary people but also scientists across diverse disciplines formulate causal hypotheses and justify them by appealing to patterns of counterfactual dependence under interventions.

Causal functionalists do not completely dismiss ontological approaches to causation. However, rather than searching for essential features that define what causation is, they advocate the "worldly infrastructure of causation" project, whose goal is to identify "certain generic features of our world that license and support the application of causal thinking and inferences to causal conclusions" (Weinberger et al., forthcoming).

4. Grand Epistemological Theories of Causation

While grand metaphysical theories of causation aim to identify the metaphysical features shared across the full range of causal relations, grand epistemological theories of causation aim to provide a

(relatively) unified methodological framework for inferring and evaluating causal hypotheses. Despite the diversity of scientific domains which causal hypotheses belong to and the diversity of scientific scenarios in which causal hypotheses are formulated, those pursuing grand epistemological theorizing about causation believe that a common methodological foundation exists for causal inference across all fields of inquiry.

For instance, philosophers and scientists who endorse evidence-based medicine or evidence-based policy regard randomized controlled trials as the "gold standard" for causal inference. Viewing causation as essentially a kind of difference-making relationship, they maintain that the kind of difference-making evidence generated by randomized controlled trials provides the strongest support for causal hypotheses, and that when such evidence is unavailable, the degree of support a causal hypothesis receives is proportional to how closely the available evidence approximates evidence generated by randomized controlled trials. If advocates of evidence-based medicine and evidence-based policy consider randomized controlled trials as the optimal method for inferring causal hypotheses not also in medicine and in policy-relevant areas but also in all other fields, they can be properly viewed as proposing a grand epistemological theory of causation that claims randomized controlled trials to be a universally valid method for causal inference.

Although grand epistemological theories of causation are often monistic, they are compatible with a modest form of evidential pluralism that identifies two or more kinds of evidence, or inference strategies, as *jointly essential* to causal inference in every domain of reality. For instance, Federica Russo and Jon Williamson maintain that both difference-making evidence and mechanistic evidence are

required to establish causal hypotheses in biomedical sciences: the former demonstrates that the putative cause does make a difference to the putative effect, while the latter explains how exactly that cause makes a difference to that effect (Russo & Williamson, 2007, see also Illari, 2011; Wilde & Parkkinen, 2019). In principle, this so-called "Russo-Williamson thesis" can be generalized to other scientific domains such as social sciences (Shan & Williamson, 2021; 2023; Ghiara, 2022, cf. Claveau, 2012) and hence be taken as a grand epistemological theory of causation.

However, the possibility of developing successful grand epistemological theories of causation is challenged by *evidential contextualism*, a more radical form of evidential pluralism that holds that the appropriate evidential bases and inferential strategies for inferring and evaluating causal hypotheses vary across scientific domains and across contexts of inquiry (Cartwright, 2006; Suárez, 2014; Shan & Williamson, 2022). Even if the "Russo-Williamson" thesis rightly claims that mechanistic evidence is required to establish causal claims in biomedical sciences, many social scientists, especially those working in the quantitative tradition, view mechanistic evidence as merely supplemental, rather than indispensable, for causal inference in social sciences. They believe that high-quality difference-making evidence alone, generated by rigorous inference strategies such as randomized controlled trials, instrumental variables, and difference-in-differences, is sufficient to warrant causal hypotheses in social sciences. Alternatively, one may treat mechanistic evidence as unnecessary for inferring that the causal hypothesis "*X* causes *Y*" holds in the population under study, yet regard it as necessary when the aim is to apply this hypothesis to a new population for some intervention purpose. If the evidential and methodological grounds for causal hypotheses are indeed context-dependent in either of these ways, it is

hard to see how any grand epistemological account of causation could correctly offer a unified methodological framework for causal inference.

5. Concluding Remarks

Grand causal theorizing is the dominant approach to theorizing about causation. Nevertheless, its success depends on the existence of generally shared features of causation in the relevant respects, which, as suggested by competing views to grand theories of causation, is not something that can be taken for granted. To vindicate the feasibility and importance of grand causal theorizing, its proponents, especially those who advocate objectivistic grand metaphysical theorizing about causation, need to engage more actively in meta-philosophical defenses of their approach.

Part II: Kinds of Causal Diversity and Middle-Range Theories of Causation

1. Introduction

For several decades, scientists have recognized that a person's level of high-density lipoprotein (HDL) cholesterol (a.k.a. the "good" cholesterol) can influence the risk of cardiovascular disease. In the 1970s and 1980s, a series of studies provided not only robust statistical evidence for an inverse association

between HDL cholesterol levels and the incidence of coronary heart disease, but also mechanistic evidence explaining how HDL cholesterol may protect the cardiovascular system and reduce the risk of cardiovascular disease (Castelli et al., 1977; Jansen & Hülsmann, 1980; Rothblat & Phillips, 1982; Tall, 1986; Abbott et al., 1988; Brown et al., 1989; Gordon et al., 1989). Taken together, this body of evidence was widely regarded as persuasively establishing a (negative) causal connection between HDL cholesterol and cardiovascular disease. Moreover, it suggested that this causal connection is *monotonic*: other things being equal, an increase in the level of HDL cholesterol will always result in either a decrease or no change in the risk of cardiovascular disease.

However, emerging evidence suggests that this causal relationship is in fact non-monotonic (Madsen et al., 2017; Li et al., 2019; Liu et al., 2022a; 2022b). While a high level of HDL cholesterol (60~80 mg/dL) can lower one's risk of cardiovascular disease, it is recently discovered that a very high level of HDL cholesterol (80 mg/dL or above) may instead increase this risk.

When scientists investigate causal relationships, they are interested not only in *whether* X causes Y but also in the *specific features* that X's causal influence on Y exhibits, since causal information about the latter is not less practically significant than causal information about the former. For example, knowing that HDL cholesterol has a non-monotonic causal influence on cardiovascular disease has clear implications for clinical practice: raising HDL cholesterol may be an effective means of preventing cardiovascular disease for individuals with HDL cholesterol below 60 mg/dL, but it may be detrimental to those with HDL cholesterol above 80 mg/dL. Here, (non-)monotonicity is *not* an essential feature of causation, i.e., a feature that defines what causal relationships are and distinguishes them from non-

causal ones. Nevertheless, it remains an important feature that is possessed by some causal relationships and distinguishes them from others.

Whether the causal relationship between *X* and *Y* (or, *X*'s causal influence on *Y*) is monotonic or not is one dimension along which causal relationships can vary; and there are many other dimensions as well. Despite the scientific and practical significance of these dimensions, they have received little attention in the philosophical literature on causation, where grand causal theorizing is the dominant approach. This is because grand causal theorizing targets primarily features that causal relationships generally share, rather than features that underwrite distinctions among different causal relationships. Consequently, important philosophical questions specific to monotonic causal relationships have received little attention. These questions may include:

- **Conceptual questions**: What does it mean for a causal relationship to be monotonic? What conceptual distinctions can be made among monotonic causal relationships?
- **Metaphysical questions**: Is monotonicity an objective feature of certain causal relationships, or is it primarily a feature of our representations or models of these causal relationships? What conditions must a causal relationship satisfy to be monotonic? Does *X*'s having a monotonic causal influence on *Y* require that the influence be mediated by one and the same mechanism throughout the full range of *X*'s possible values?

- **Epistemological questions**: What constitutes an adequate evidential basis for scientists to judge that a causal relationship is monotonic?
- **Pragmatic questions**: When there is only partial evidence for the monotonicity of a causal relationship, what is the best practice for using this causal relationship for pragmatic purposes such as prediction and intervention?

To address these questions, we need to shift from grand causal theorizing to a more nuanced theorizing approach to causation.

Part II of this chapter will focus on *middle-range causal theorizing*, a theorizing approach that primarily targets the diverse and heterogeneous respects of causation, in contrast to grand causal theorizing, which primarily targets its unified and homogeneous respects. I will identify four important respects in which causation is diverse: causal relationships, causal systems, causal domains, and causal roles (Section 2). Recognizing these kinds of causal diversity as well as their scientific significance motivates the development of philosophical theories of causation that are middle-range in character. I will introduce the idea of middle-range theories of causation and distinguish three major variants: feature-centered theories, domain-centered theories, and role-centered theories of causation (Section 3).

2. Kinds of Causal Diversity

The increasing popularity of the term "causal pluralism" over the past two decades indicates a growing recognition that causation is diverse in some respects. The question is: in what respects? There

are at least six kinds of causal diversity that can be identified from the relevant philosophical and scientific literature:

(a) The diversity of the nature (or essential features) of causal relationships (ontological pluralism);

(b) The diversity of the evidential/methodological grounds for causal inference (evidential pluralism);

(c) The diversity of the non-essential features of causal relationships;

(d) The diversity of causal systems;

(e) The diversity of causal domains;

(f) The diversity of causal roles.

The first two kinds of causal diversity have received more attention than others in the literature. They are also briefly covered in Part I. For these reasons, the rest of this section will focus on the other four kinds of causal diversity. Meanwhile, it is worth emphasizing that there are competing views about how, and to what extent, the essential features of causation and the evidential grounds for causal inference are diverse. As also noted in Part I, whereas modest forms of ontological pluralism and of evidential pluralism can be accommodated by grand causal theorizing (e.g., Hall, 2004; Russo & Williamson, 2007), more radical forms of these views, such as causal particularism and evidential contextualism, deny the prospect of grand causal theorizing in the relevant respects.

2.1 The Diversity of the Non-Essential Features of Causal Relationships

A causal relationship possesses not only essential features but also non-essential ones. A feature of a causal relationship is non-essential if and only if it is not among the set of features that jointly define what causation is, that is, features whose absence would render the relationship not causal.[5] While individual causal relationships may share the same set of essential features (if monism about the nature of causation is true), they vary significantly in what non-essential features they possess or exhibit. By examining which features scientists attend to in actual causal research, we can identify a repository of non-essential yet scientifically and practically important features of causation. Here is an incomplete list of the non-essential features that a causal relationship between X and Y may exhibit:[6]

- Monotonicity: An increase/decrease in X always causes an increase/decrease or no change in Y. Otherwise, a causal relationship is non-monotonic (VanderWeele & Robins, 2009; Liu et al., 2022a; 2022b).
- Linearity: The causal influence of X on Y remains the same across different values of X. Otherwise, a causal relationship is non-linear (Gourévitch et al., 2006; Mulaik, 2009; Fernandez, 2014).

[5] Ross (manuscript) refers to essential features and non-essential features as "primary features" and "secondary features" respectively.

[6] Most non-essential features of causal relationships listed here are adapted from Gerring (2011).

- Reversibility: *X*'s causal influence on *Y* can be reversed by returning *X* to its original state. Otherwise, a causal relationship is irreversible (Ross & Woodward, 2022; Stephan, 2025).
- Constancy: *X* exists constantly and operates continuously upon *Y* (Caudill, 1973; Antecol, 2000).
- Delimitedness: *X* operates only briefly on *Y* (though its influence on *Y* may endure) (Karbownik & Wray, 2019; Yagan, 2019).
- Proximality: *X* has an immediate causal influence on *Y* (Mayr, 1961; Laland et al., 2011; Shan, 2025).
- Distality: *X* takes a considerable amount of time to influence *Y* (ibid).
- Threshold effect: *X* can have a causal influence on *Y* only when it exceeds a particular threshold level (Jauk et al., 2013; Tran et al., 2022).
- Path-dependency: *X* has an enduring, and perhaps increasing or self-reinforcing, influence on *Y* over time (David, 1985; Pierson, 2000a; Berger, 2009).
- Set-theoretic sufficiency: *X* is a sufficient condition for *Y*. That is, there is no actual case where *X* occurs, but *Y* does not (Skocpol, 1979).
- Set-theoretic necessity: *X* is a necessary condition for *Y*. That is, there is no actual case where *Y* occurs, but *X* does not (Skocpol, 1979).

- Determinacy: When *X* occurs, *Y* occurs without exception under normal background conditions. Otherwise, a causal relationship is non-deterministic (Feder et al., 2012).
- Stability: *X* remains to have a causal influence on *Y* under changes in background conditions (Woodward, 2010).
- Specificity: *X* influences and only influences *Y* (ibid).
- Proportionality: *X* is at the "right" level relative to *Y* such that changes in the state of *X* are associated with changes in the state of *Y* (ibid).
- .....

We may regard these non-essential features as capturing different possible *modes* of *X*'s causal influence on *Y*.

Two points to note. First, all of these non-essential features of causal relationships (or, modes of causal influence) are *domain-general*, which means that they are characterized in a way that does not conceptually preclude causal relationships in any particular domain of reality from exhibiting them. Thus, when a non-essential feature is found to be rarely, or even never, exhibited by causal relationships in a particular domain, it is worth asking what features of the domain account for the rarity or absence of this feature.

Second, some non-essential features of causation can be exhibited only by causal relationships that satisfy certain representational requirements, whereas some others do not appear to impose such

requirements. For example, for education level to stand in a linear causal relationship with socioeconomic status, both causal factors must be meaningfully represented as numerical variables, even if their original states are not inherently numerical. By contrast, whether a causal relationship is reversible does not seem to depend on how it is represented or modelled. The fact that some non-essential features are representation-dependent while others are not suggests that not all non-essential features are objective features of causal relationships to the same extent.

2.2 The Diversity of Causal Systems

A causal system is a structured arrangement of causal relationships among multiple causal factors. Causal systems come in many varieties, differing both in the kinds of causal factors they involve and in the structural features that characterize how those factors are related to one another. Some varieties (e.g., causal systems exhibiting structural features like preemption, overdetermination, or double prevention) have attracted more philosophical attention because they pose challenges to prominent grand metaphysical theories of causation. Still, many other scientifically significant varieties of causal system remain relatively underexamined in philosophy.

One major exception is mechanisms, which have been extensively examined over the last three decades (Glennan, 1996; 2017; Machamer et al., 2000; Woodward, 2002; Bechtel & Abrahamsen, 2005; Craver, 2007; Illari & Williamson, 2012; Ioannidis & Psillos, 2022). According to a minimal definition, "[a] mechanism for a phenomenon consists of entities (or parts) whose activities and interactions are organized so as to be responsible for the phenomenon" (Glennan, 2017: 17; see also Illari & Williamson,

2012). This definition suggests that mechanisms are typically hierarchical: the higher-level behavior of the mechanism as a whole is constituted by the lower-level activities and interactions of its parts. Although the notion of mechanism captures a wide range of causal systems in science, Lauren Ross (2026) recently argues that mechanisms need to be distinguished from another kind of complex causal system that scientists refer to as "cascades." Cascades involve (i) an initial trigger that "kick-starts" subsequent causal steps, (ii) a sequence of amplifying steps that cumulatively produces a much larger effect compared to the initial trigger, and (iii) stable progression from start to finish. A crucial difference between cascades and mechanisms is that cascades are not typically hierarchical in the way mechanisms are.

Other kinds of causal system that deserve more philosophical scrutiny due to their scientific significance include those exhibiting the following structural features:

- Causal chain: A sequence of causal relationships links an initial cause to an ultimate effect (Baumgartner, 2008; Gross, 2018).
- Equifinality: The same effect can be produced by different independent causes through different causal pathways (Cicchetti & Rogosch, 1996; Stanley & Sawyer, 2009).[7]

[7] The structural pattern of equifinality is similar to that of overdetermination. Here I conceptually distinguish them in order to highlight the fact that equifinality needs more philosophical inquiry to set it in a scientific context, rather than being treated merely as a philosophical intriguing case.

- Multifinality: The same cause can lead to different effects depending on the context (Cicchetti & Rogosch, 1996; Nolen-Hoeksema & Watkins, 2011).
- Interaction effect: The interaction between two or more independent causes of an effect produces a novel causal influence on the effect that cannot be reduced to the mere sum of their individual causal influences (Hargens, 2009; Egami & Imai; 2019).
- Sequence effect: The causal influence of causes $X_1, X_2, ..., X_i$ on Y depends upon the timing and order in which $X_1, X_2, ..., X_i$ occur (Pierson, 2000b).
- Causal cycle (or feedback loop): X has a causal influence on Y, which in turn has a subsequent causal influence on X (Warrell & Gerstein, 2020; Bechtel & Bollhagen, 2024).
- .....

Similar to non-essential features of causal relationships, these structural features of causal systems are characterized in a domain-general way, which permits them to be exhibited by causal systems across different scientific domains.

2.3 The Diversity of Causal Domains

Science is naturally divided into different scientific domains such as physics, biology, psychology, and economics (as well as into more fine-grained subdomains such as quantum physics and behavioral economics). These scientific domains investigate causation in their respective domains of reality. I will

use "causal domains" to refer interchangeably to scientific domains and to corresponding domains of reality they study. Because each causal domain involves its own set of entities whose distinctive characteristics can shape the causal relationships or causal systems they enter into, causation in a causal domain often exhibit *domain-specific* features that distinguish it from causation in other domains.

Domain-specific causal features may come from three main sources. First, different causal domains often exhibit "representative" features that are characteristic of the causal relationships or causal systems in those domains. For example, universal causal laws are widely taken to be pervasive in physical sciences but are rarely found in biological and social sciences. Instead, causal relationships in the latter domains are often understood as grounded in local causal mechanisms that operate within specific causal contexts (Machamer et al., 2000; Hedström, 2005; Falleti & Lynch, 2009; Hedström & Ylikoski, 2010; Craver & Darden, 2013).

Second, causation in different causal domains can exhibit the same domain-general feature in distinct domain-specific ways. Although mechanisms are prominent in both biology and sociology, biological mechanisms and social mechanisms presumably differ in certain important characteristics. This is partly because entities that play causal roles in biological mechanisms (e.g. cells, genes, organisms) tend to have causally relevant characteristics that differ from those of the entities that play causal roles in social mechanisms (e.g. humans, institutes, social structures).

Third, some causal features are unique to, though not necessarily typical of, causal relationships in a given domain. As an example, ideational causation concerns cases in which one agent's ideas (e.g.,

pacifist beliefs and values) influence their own or others' decisions (e.g., the decision to negotiate a ceasefire) or behavior (e.g., participation in an anti-war protest) (Yee, 1996; Jacobs, 2014). Whether or not ideational causation is widespread in the social domain, it is unique to this causal domain because only social agents can have their mental states affect other actors through verbal communication and other forms of social interaction. In this respect, ideational causation exhibits a distinctive domain-specific feature.

2.4 The Diversity of Causal Roles

Causal relationships are often characterized as relationships between causes and effects. However, this characterization may conceal the diversity of causal roles that causal factors play relative to an effect. Here are some examples:

(1) The lightning *causes* a forest fire.

(2) The presence of oxygen *enables* the lightning to cause a forest fire.

(3) Forest rangers' timely intervention *prevents* the lightning from causing a forest fire.

(4) High temperatures *hasten* the onset of a forest fire.

(5) A sudden rain *delays* the onset of a forest fire.

(6) Strong winds *intensify* the forest fire.

(7) High humidity *diminishes* the intensity of the forest fire.

Each of these causal factors plays a distinct role relative to the effect in question. In general, relative to a causal relationship in which C causes E, some causal factors may function as *enablers* (or enabling

conditions), which allow *C* to bring about *E* in a given scenario. Others may function as *derailers* (or disabling conditions), which prevent *C* from bringing about *E* in a given scenario. Besides these on-off conditions, there are also various kinds of "effect modifier" or "moderator" that can modulate how *C* actually affects *E*. For instance, some causal factors may be *hasteners*, which make *E*, as a consequence of *C*, occur earlier than it otherwise would, whereas others may be *delayers*, making *E* occur later than it otherwise would. Some may be *intensifiers*, which increase the magnitude of the causal influence of *C* on *E*, whereas others may be *diminishers*, which decrease the magnitude of the causal influence of *C* on *E*.

Identifying diverse causal roles can help us better articulate the notion of "background condition" (or "causal context"). It is widely agreed that a cause rarely brings about an effect on its own. When a cause actually produces its effect in a given scenario, it typically does so only relative to a causal context that consists of the relevant background conditions present in that scenario. However, for a long time, background conditions have been understood as consisting merely of enablers, despite the fact that factors playing other causal roles may also participate in the causal process through which a cause produces its effect. A more adequate understanding of background conditions needs to encompass not only enablers but also relevant factors that play other causal roles in that scenario.

Furthermore, the recognition that there are multiple causal roles distinct from the role of being a cause raises the question of whether the former are in fact dependent on, or reducible to, the latter. That is, for each causal role *R* distinct from being a cause, one can ask: when a factor *F* plays causal role *R* relative to the causal relationship between *C* and *E*, is *F* thereby itself a cause of *E*, or is it causally

related to *E* in a way that cannot be reduced to causation? Existing philosophical discussions of this question have focused primarily on the relationship between enablers and causes. Some defend the "causal parity" thesis—the view that there is no intrinsic difference between causal factors that function as causes and those that function as enablers (Mill, 1874; Mackie, 1974; Lewis, 1986b), whereas others reject it (Waters, 2007; Broadbent, 2008; Ross, 2018). Similar discussions also exist regarding whether hasteners and delayers can be properly viewed as causes (Mackie, 1992; Sartorio, 2006; Touborg, 2018). Nevertheless, other causal roles remain underexplored on this issue.

3. Middle-Range Theories of Causation

Middle-range theories of causation apply the notion of middle-range theorizing to the philosophical study of causation. This notion was originally articulated by the sociologist Robert K. Merton (1949) as an alternative to the form of all-encompassing and highly abstract theorizing (a.k.a. "grand theorizing") that dominated his time. Middle-range theories are "middle" with respect to the level both of generality and abstraction (Cartwright, 2020). Whereas grand sociological theories aspire to explain all social phenomena within a single unified framework, middle-range theories are "semigeneral theories each of which is adequate for explaining a limited range or type of phenomena" (Hedström & Udehn, 2009: 31). This narrower theoretical scope corresponds to a lower level of abstraction: while grand theories typically abstract away from the distinctive features of particular social phenomena, middle-range theories tend to preserve information about these features, thereby enabling analysis of specific causal mechanisms underlying their target phenomena and yielding a more accurate and fine-

grained understanding. Nowadays, middle-range theories are ubiquitous across social sciences and life sciences.

A parallel distinction can be drawn among philosophical theories of causation (as well as philosophical theories on other topics). Grand theories of causation are highly general insofar as they aim to account for all causal relationships, and highly abstract insofar as they abstract away from a significant number of details about particular causal relationships, preserving only what is presumed to be widely shared across causal relationships. This approach is well-suited to philosophical projects that seek genuinely shared features of causation. Nevertheless, its high level of abstraction precludes it from illuminating the distinctive features of diverse kinds of causal relationship, whose understanding is also of paramount scientific and practical significance. For this reason, grand theories of causation on their own cannot yield an adequate philosophical understanding of causation (*the Inadequacy Thesis*).

In contrast to grand theories of causation, middle-range theories of causation operate at an intermediate level of generality and abstraction. Each middle-range theory targets a specific kind of causal relationship and preserves information distinctive to it. By addressing philosophical questions specific to its target kind of causal relationship, a middle-range theory of causation can thus yield a more fine-grained understanding of the causal relationships within its theoretical scope. Taken together, middle-range theories of causation provide a means to shed light on the sheer diversity of causal relationships, which unification-seeking grand theories of causation are unable to do.

Echoing the varieties of causal diversity, a middle-range theory of causation can be either (a) feature-centered, (b) domain-centered, or (c) role-centered. A *feature-centered* middle-range theory of causation focuses on causal relationships that exhibit a certain non-essential feature or on causal systems that exhibit a certain structural feature. It typically begins by identifying a domain-general causal feature and proceeds to offer philosophical (e.g. conceptual, ontological, epistemological, or pragmatic) insights applied to causal relationships or causal systems with this feature. For instance, Ross and Woodward (2022) develop an account for irreversible causal relationships, in which they conceptually clarify what it means for causal relationships to be irreversible and discuss the implications of irreversible causation for causal inference and modelling.

Nevertheless, existing attempts to develop feature-centered accounts of causation fall short in two respects. First, while certain causal features of causal relationships or causal systems—such as stability, specificity, proportionality, and causal cycles—have recently been subjected to philosophical scrutiny, many non-essential causal features remain underexplored. For instance, little philosophical work has been devoted to causal relationships that are path-dependent or monotonic. Second, even in cases where particular causal features have received philosophical attention, they are rarely examined from a metaphysical perspective. This leaves their metaphysical status obscure. For example, although some conceptual and epistemological issues concerning (ir)reversible causal relationships have been analyzed, we still lack an account of the metaphysical grounds that explain why some causal relations are reversible whereas others are not.

By contrast, a *domain-centered* middle-range theory of causation, such as a theory of causation in economics (Hoover, 2001; Child, 2020; Maziarz, 2020) or a theory of genetic causation (Waters, 2007; Stegmann, 2014), focuses on causal relationships and causal systems within a particular causal domain (or sub-domain), and seeks to address philosophical questions specific to causation in that area. As noted earlier, the distinctive set of entities studied in each causal domain often shape the relevant features of the causal relationships and causal systems they participate in. Consequently, many philosophical questions about causation that arise within a causal domain require domain-specific treatments that may not be generalizable to other causal domains. A domain-centered middle-range theory can thus provide philosophical insights tailored to causation in a given causal domain by, for example, identifying *representative* or *unique* causal features in that domain and addressing the conceptual, ontological, epistemological, or pragmatic issues these features raise.

Domain-centered accounts of causation have been developed for many major causal domains. However, some domains (e.g., physics, biology, economics) have received considerably more philosophical attention than others (e.g., chemistry, political science). This has resulted in an unevenly distributed philosophical understanding of causation across major causal domains. Moreover, causation in many sub-domains has not yet been adequately examined. For instance, there is as yet no philosophical account of ideational causation that systematically illuminates its conceptual, metaphysical, epistemological, and pragmatic features, despite its causal significance in both everyday life and scientific practice.

Distinguished from both feature-centered and domain-centered middle-range theories of causation, a *role-centered* middle-range theory of causation focuses on a specific causal role that a causal factor may play in relation to a causal relationship, and addresses philosophical questions related to this causal role. For instance, a theory of hasteners may aim to accomplish tasks such as defining what it means for a causal factor to hasten the effect of a cause, analyzing the relationship between a hastener and the cause(s) of the effect it hastens, offering a methodological guide for establishing claims such as "Factor *F* is a hastener of effect *E* caused by cause *C*," and/or highlighting the importance of identifying hasteners in practical contexts such as medical treatment or policy making. Nevertheless, existing role-centered accounts have focused primarily on enablers, leaving the nature of causal factors that serve other causal roles underexplored.

4. Concluding Remarks

Grand causal theorizing and middle-range causal theorizing are two complementary approaches to theorizing about causation. While the former sheds light on the features that unify causation, the latter illuminates the distinctive characteristics of specific kinds of causal relationship. Both approaches are valuable for a well-grounded philosophical understanding of causation. In the worst-case scenario, even if no successful grand metaphysical or epistemological theory of causation could be developed in light of the challenges such theories face, middle-range theorizing would remain a robust and fruitful source of philosophical understanding of causation.

In addition to their intrinsic importance, middle-range theories of causation can also contribute to the development of successful grand theories of causation. By carefully and thoroughly examining important varieties of causation, middle-range causal theorizing can provide a much richer set of test cases against which the plausibility of a given grand theory of causation can be assessed.

Nevertheless, in contrast to the popularity of grand theories of causation, there remains a shortage of middle-range theories of causation, given the various kinds of causal diversity suggested by scientific research across different domains. Although many philosophers of science and scientists have engaged in the development of domain-centered middle-range theories of causation, many causal domains and sub-domains are still underexplored from a philosophical perspective. Besides, much less attention has been devoted to the development of feature-centered and role-centered middle-range theories. This gap warrants the need for philosophers, especially those advocating for a "naturalistic" approach, to devote more intellectual resources to developing middle-range theories of causation.

Acknowledgements

Chapter 1, in full, is a reprint of the material currently under review at *Philosophy Compass*.

Chapter 2

Facilitated Causation

1. Introduction

Enablers (or enabling conditions) are background conditions whose presence enables, or makes it possible for, a cause to produce its effect (Mackie, 1974; Lombard, 1990; Cheng & Novick, 1991; Yablo, 2004). Suppose that my act of scrubbing a greasy plate yesterday caused the grease to be removed from the plate. The relevant enablers for this (token) causal relation include my use of a sponge and the presence of hot water, because they made it possible for my scrubbing to trigger a causal process that culminated in the complete removal of the grease. If I had not used a sponge or there was no hot water, my scrubbing would not have caused grease removal.

Two main questions concerning enablers arise. First, how should enablers be defined? The previous example suggests a simple counterfactual account of enabling. For events *b*, *c*, and *e* that occur in some situation *S*,

(E1) *b* enables *c* to cause *e* if and only if (hereafter "iff"), had *b* not occurred, *c* would not have caused *e*.[8]

By "event," I include not only dynamic events, such as my scrubbing a plate or John's striking a match, but also static events (i.e., states of affairs), such as the presence of hot water or of oxygen. By "enabler," I refer only to events that *actually* enable a cause to produce its effect in a given situation, and not to events that merely have the capacity to do so without actually exercising that capacity (though we may denote such events as "potential enablers"). On this account, enablers are characterized essentially as difference-makers with respect to *whether* a cause produces its effect: the absence of an enabler, which exercises its enabling capacity in an actual situation, would prevent the cause from producing the effect in that situation.

Second, are enablers intrinsically distinct from "foregrounded" causes? That is, when multiple events cooperate to produce an outcome, are there objective criteria for foregrounding some of them as the cause(s) and backgrounding the others as mere enablers? This is the so-called problem of causal selection. Three views emerge in the literature:

(**Subjectivism**) Both foregrounded causes and enablers are causes in the proper sense. The boundary between them is solely determined by subjective or pragmatic

[8] My discussion focuses primarily on singular (or, token-level) causation, i.e., causal relations between events, because the role of background conditions is most prominent in this context.

considerations and may vary across contexts. (Mill, 1874; Mackie, 1974; Hart & Honoré, 1985; Lewis, 1986b).

(**Moderate Objectivism**) Both foregrounded causes and enablers are causes in the proper sense. However, there is an objective boundary between them, insofar as foregrounded causes possess additional ontological features that enablers lack (Waters, 2007).

(**Strong Objectivism**) There is an objective boundary between foregrounded causes and enablers, insofar as only foregrounded causes are causes in the proper sense (Lombard, 1990).

Both subjectivists and moderate objectivists are motivated by the observation that most, if not all, mainstream accounts of causation classify the enablers for a cause to produce its effect as causes of that effect as well. They therefore endorse the claim that enablers are *proper causes*, that is, causes recognized as such by our best account of causation. A rejection of this claim will thus undermine both subjectivism and moderate objectivism.

I will address both questions in this chapter. But rather than engaging with them directly, I will take an indirect approach by focusing primarily on a particular kind of background condition that I call *facilitators* (or facilitating conditions). To get an intuitive sense of what facilitators are, consider again the situation in which I scrubbed a greasy plate, but with dish soap. As a result, the grease came off much more easily than it would have if I had not used dish soap. Here, my scrubbing the plate caused the grease

to be removed, while my use of a sponge and the presence of hot water served as the relevant enablers. But what causal role did my use of dish soap play? It did not seem to be an enabler, since its presence made no essential difference to *whether* my scrubbing caused grease removal. Instead, what it changes is *how* my scrubbing caused grease removal. More specifically, it increased the rate at which my scrubbing removed the grease, so that the grease was completely removed at an earlier time. It also made my scrubbing more efficacious, in the sense that the same amount of scrubbing effort could remove more grease within the same period of time.

In general, facilitators are background conditions whose presence does not affect whether a cause produces its effect but instead increases the efficiency with which it does so by, for example, increasing the rate of change and/or causal strength of the relevant causal process, as in the case of my use of dish soap. Consequently, in cases of "facilitated causation," where a facilitator helps a cause produce its effect more efficiently, the presence of the facilitator does not make a difference to whether the effect occurs, but instead make it occur earlier and/or with a greater magnitude. For simplicity, I will refer to the cause and the effect in a case of facilitated causation as the "facilitated cause" and the "facilitated effect" respectively.

Despite their causal significance in scientific practice and ordinary scenarios, facilitators have received little philosophical attention. In what follows, I tackle two conceptual puzzles about facilitators. One concerns their relation to enablers: are facilitators essentially distinct from enablers? The answer may seem to be yes. Enablers make it possible for a cause to produce its effect, whereas facilitators merely make a cause that is already in a position to produce its effect do so more efficiently. It is therefore

tempting to regard facilitators as a distinct kind of background condition from enablers, which suggests a substantially disunified picture of background conditions.

However, after illustrating the causal significance of facilitators in scientific contexts (Section 2), I argue that facilitators are best understood as *redundant enablers*, i.e., enablers that redundantly enable a cause to produce its effect in the presence of other enablers that are themselves sufficient to do this job. It follows that the simple counterfactual account of enabling must be rejected because it cannot accommodate cases of redundant enabling. A more adequate account construes enablers as difference-makers not for causal relations between causes and effects, but for the causal processes that realize those relations (Section 3). This modified account of enabling, I argue, implies that facilitators are redundant enablers (Section 4). Recognizing the causal structures of redundant enabling in turn supports a more unified picture of background conditions, on which the apparently diverse causal roles played by a wide range of background conditions are reducible to a single causal role: enabling (Section 5).

Another puzzle concerns whether facilitators are proper causes. When a facilitator helps a cause to produce its effect more rapidly or more efficaciously, is it also a cause of the facilitated effect in the proper sense? I argue that, unlike most enablers, facilitators are not classified by mainstream difference-making accounts of causation as causes of the facilitated effects. Given that facilitators are redundant enablers, this failure provides counterexamples to the widely held thesis that all enablers are proper causes. The falsity of that thesis undermines not only subjectivism and moderate objectivism about causal selection but also the view that causation is the only fundamental kind of causal relevance (Section 6). I then show that a more promising way to defend facilitators as genuine causes, while preserving the idea

that causes make a difference to their effects, is to endorse a hybrid account that combines the difference-making idea with an emphasis on causal processes, which has traditionally been central to mechanistic accounts of causation. On this account, causes are difference-makers for the causal processes that lead to their effects (Section 7).

2. The Causal Significance of Facilitators

Throughout the chapter, I assume a relatively coarse-grained individuation of events, according to which a significant class of properties are accidental to events (Lewis, 1986c; Strevens, 2004). I further assume that the accidental properties of an event include its time, location, and manner of occurrence, unless there are good reasons to individuate a specific event in terms of one or more of these properties (see also Sartorio, 2006). Thus, an event, by default, can vary with respect to its timing, location, magnitude, and so on while still being regarded as the same event.

In causal investigations, scientists are often interested in not only figuring out whether an antecedent event causes a given outcome but also inferring key features of causal influence that characterize how an identified cause affects its outcome. These causal features help account for why the outcome has certain accidental properties. For example, to explain why an event occurs at a particular time, it is useful to know when its putative causes occurred and at what rate they produced the event. Likewise, to explain why an effect occurs with a particular magnitude, it is useful to know the size of its cause and the causal strength of that cause with respect to the effect.

Causal features that characterize how a cause influences another are often subject to adjustment, because other happenings may shape the causal process connecting a cause to its effect by interacting with either the cause itself or the mediating events it produces. As a result, some factors (*accelerators*) may speed up the relevant causal process so that a shorter amount of time is needed for a cause to produce its effect, thereby making the effect occur earlier. Other factors (*decelerators*) may slow down the relevant causal process and thus delay the effect. A good understanding of rate-adjusting factors can improve our ability to predict, explain, and manipulate the timing of an expected outcome. For instance, during the COVID-19 pandemic, many countries implemented large-scale anti-contagion policies (e.g., social distancing, mask mandates) to delay the point at which new infections peaked, thereby giving regional health systems additional time to expand hospital capacity and vaccine coverage (Hsiang et al., 2020; Adjodah et al., 2021). Granted that a country's epidemic peak is an expectable consequence of the COVID-19 outbreak in that country, anti-contagion policies can be viewed as effective decelerators. These policies did not directly generate or eliminate infections, but instead significantly slowed the growth rate of new infections and thus postponed the epidemic peak in a country.

There are also factors whose presence may affect the causal strength of a cause for a target effect, in cases where the effect involves certain features that can vary in their magnitude. Some factors (*amplifiers*) may magnify or expand the impact of a cause of a given magnitude, which allows it to produce an effect of greater magnitude. Other factors (*diminishers*) do the opposite. They may mitigate or constrain the impact of a cause and thus reduce the magnitude or scope of its effect. Strength-adjusting factors are also of scientific interest, since they help scientists better predict and explain why an effect has

a certain magnitude, and enable them to enhance desirable effects or mitigate undesirable ones through intervention.[9] For instance, chronic hypertension can lead to intracerebral hemorrhage by causing structural damage to cerebral small vessels. Patients receiving oral anticoagulants may develop more severe intracerebral hemorrhage because anticoagulation impairs the blood's ability to clot. Clinicians can therefore improve prediction of severe intracerebral hemorrhage in hypertensive patients by taking anticoagulation status into account (Flaherty et al., 2008; Veltkamp et al., 2013).

Facilitators include both accelerators and amplifiers. These categories capture two major ways in which the presence of a factor can enhance the efficiency with which a cause produces a target effect. Note that accelerators and amplifiers should not be considered as totally distinct, non-overlapping categories. In many cases, the same factor can both accelerate the rate at which a cause produces its effect and amplify its causal strength with respect to that effect. This is perhaps because speeding up a causal process allows the cause to produce a larger effect within the same period of time (e.g., strong wind increasing the rate at which a spark causes a fire to spread, thereby resulting in a larger burned area within the same period), or because greater causal strength allows the cause to bring about the same effect at an

[9] Rather than using terms such as "amplifiers" and "diminishers," scientists often refer to strength-adjusting factors as "moderators." Note, however, that not all moderators can properly be regarded as amplifiers or diminishers. Some moderators make a difference to the causal strength of a causal relation at the level of general causation, but not at the level of singular causation. For instance, sex may be regarded as a moderator of the effectiveness of a medical treatment for some disease if the treatment has different average effects on male and female patients. However, for any individual patient who receives the treatment, it is impossible to alter the treatment's effectiveness by adjusting that patient's sex. By contrast, the strength-adjusting factors on which I focus can adjust causal strength in singular causal relations.

earlier time (e.g., the presence of dish soap enhancing the causal strength of my scrubbing with respect to removing grease, thereby allowing the grease to be completely removed sooner).

The causal significance of facilitators is widely recognized in both scientific inquiry and everyday causal thinking. One major context in which facilitators play a significant role is chemical reactions. Chemical reactions may occur spontaneously under suitable thermodynamic conditions. A chemical reaction can be plausibly understood as a causal relation in which certain reactants produce certain products, while the thermodynamic conditions (e.g., temperature, pressure) that trigger the reaction serve as enablers for this causal relation to obtain. However, even when the required thermodynamic conditions are in place, many reactions still proceed relatively slowly because reactant molecules must overcome a high activation energy in order to break existing chemical bonds and initiate the reaction. A central task in chemistry, therefore, is to identify suitable catalysts for chemical reactions of interest, as their presence can significantly accelerate relevant reactions by lowering the activation energy required for them to proceed.

An adequate account of causation in chemistry needs to account for the causal role of catalysts (Seifert, 2024). Yet most philosophical discussions of chemical causation have neglected catalysts (Goodwin, 2012; Harré, 2016; Hendry, 2023). As a rare exception, Lauren Ross (2018) argues that enzymes, which are catalysts in biochemical reactions, are often selected as causes of the formation of final products, together with the metabolic substrates from which those products are ultimately formed. This is because enzymes exert an appropriate kind of causal control over whether the products are formed within a timescale relevant to human interests. However, Ross's account cannot capture the role

of catalysts in cases where a chemical reaction would already proceed on a relevant timescale without the presence of any catalyst. In such cases, the introduction of a catalyst can still increase the reaction rate, even though it makes no difference to whether the product is formed within a suitable timescale. A more promising starting point, in my view, is to understand catalysts as facilitators that help reactants form products more efficiently, that is, at a much faster rate and with greater causal strength. Then we may investigate whether catalysts also count as causes in both cases where the presence of a catalyst makes a substantial difference to whether a reaction occurs and cases where its presence merely accelerates the reaction without making a difference to whether it occurs.

In fact, the term "catalyst" has been widely used in causal discourse that goes beyond chemical contexts as a metaphor for causal factors whose presence significantly accelerates the occurrence of a given outcome. For example, social scientists have characterized a number of factors including lawyers (Moliterno, 2009), social media platforms (Yeung, 2018), and justice-centered science pedagogy in high schools (Morales-Doyle, 2017) as catalysts for social change, in the sense that their presence can facilitate its occurrence.

Moreover, causal factors described as catalysts for an outcome are often deliberately distinguished from its putative causes. For instance, social psychologists recently argue that, given the lack of robust meta-analytic evidence for a statistically significant and consistent causal effect of economic inequality on mental health, economic inequality "may be better understood as a *catalyst* for other determinants of health (for example, poverty and inflation), rather than as a *root cause* of diminished subjective well-being or mental health problems" (Sommet et al. 2026: 935; emphasis added).

The direct contrast between causes and catalysts (or facilitators) thus raises a conceptual puzzle: should a catalyst for a cause to bring about an outcome also be seen as a cause of that outcome? This question will be addressed in Sections 6 and 7.

3. The Problem of Redundant Enabling

What is the nature of facilitators? In particular, how are they related to more familiar kinds of background conditions such as enablers? In the next section, I will argue that facilitators are not a *sui generis* kind of background condition, but rather belong to a special kind of enablers, i.e., redundant enablers. To accurately assess whether facilitators are in fact enablers, a plausible account of enabling is needed. This is my objective in this section.

The simple counterfactual account of enabling, E1, characterizes enablers as difference-makers for whether a cause produces its effect. At first pass, this account fits well with many familiar causal judgments. My use of a sponge enables my scrubbing a greasy plate to cause grease removal because if I had not used a sponge, my scrubbing would not have caused the grease to be removed. The presence of oxygen enables John's striking a match to cause its lighting because if there had not been oxygen, John's striking would not have caused the match to light. Thus, E1 seems to offer a reliable guide for whether an event is an enabler. Given E1, facilitators are not enablers because they do not make a difference to whether a cause produces its effect. If I had not used dish soap, my scrubbing would still have caused grease removal.

However, E1 is flawed because it cannot accommodate cases of redundant enabling, in which one or more enablers redundantly enable *c* to cause *e*, in the sense that even if those enablers had been absent, other enablers would still have enabled *c* to cause *e* in the same situation. Consider this example:

> **Two Soldiers**: Soldier A and Soldier B stand by, waiting for the general's order to execute a murderer. The general shouts, "Execute!" Both soldiers fire at the murderer, each resulting in a fatal wound. As a result, the murderer dies.

For the sake of argument, grant that the general's order causes the murderer's death. Two independent sets of enablers can be found in the situation. One set consists of the conditions pertaining to Soldier A, such as her being ready to respond to the order, her physical capacities required for firing at the murderer, her gun's proper functioning, and so on. In a counterfactual situation in which only Soldier A had stood by, her relevant conditions would have enabled the general's order to trigger a causal process, P_A, which involves a sequence of events including Soldier A's firing, her bullet travelling a particular trajectory, and its eventually hitting the murderer, before reaching the outcome that the murderer is killed.

The other set of enablers consists of similar conditions pertaining to Soldier B. In a counterfactual situation in which only Soldier B had stood by, her relevant conditions would have enabled the general's order to trigger a distinct causal process, P_B, which involves a sequence of events including Soldier B's firing, her bullet travelling another particular trajectory, and its eventually hitting the murderer, before reaching the outcome that the murderer is killed.

Figure 2.1 represents the causal structure of *Two Soldiers*, where both Soldier A and Soldier B stand by. Before getting into its details, some general remarks are needed. Standard causal diagrams used by philosophers, such as neuron diagrams and causal graphs, represent causation as a binary relation between a cause and an effect. This choice makes them unable to represent enablers faithfully, since it allows enablers to appear only as additional causes of the effect alongside the cause they enable. Consequently, standard causal diagrams cannot distinguish between enabled causation, in which a cause produces an effect with the support of relevant enablers (for example, the presence of hot water enables my scrubbing a plate to causes the grease to be removed), and causal overdetermination, in which multiple independent causes simultaneously produce the same effect (for example, John's smoking habit and air pollution independently cause his lung cancer).

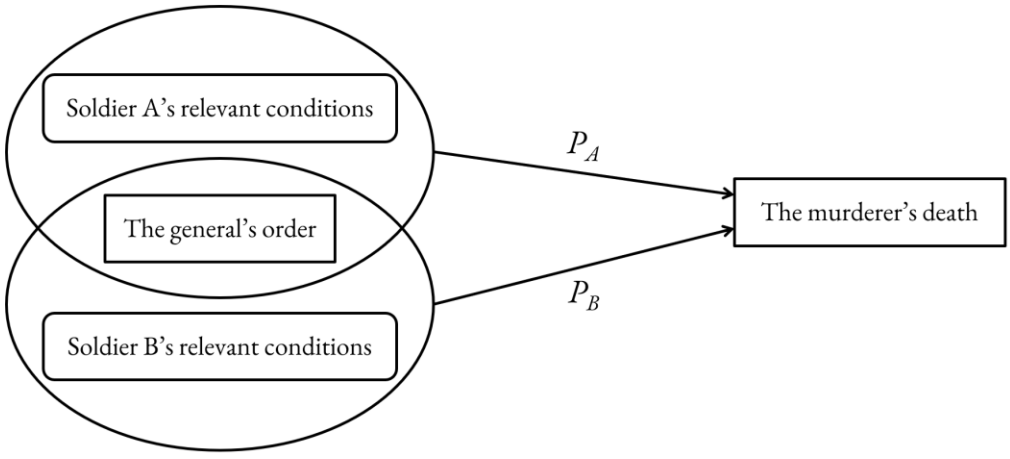


Figure 2.1 A case of over-enabling.

One ready graphical tool for faithfully representing enabled causation is the enriched Theory of Change (eToC) diagram developed in Cartwright et al. (2026), of which Figure 2.1 is an instance. An eToC diagram represents a causal relation—or, more precisely, a causal process—as holding between, on the one hand, a bundle that consists of the cause and its relevant enablers, and, on the other hand, the effect. Standard rectangles represent core events such as causes and effects. Rounded rectangles represent enablers. Ellipses represent cause-enabler bundles. And bold lines represent causal relations (or causal processes) that obtain in the situation.[10]

As depicted in Figure 2.1, in the actual situation where Soldier A and Soldier B stand by, both of their relevant conditions actually play an enabling role. Soldier A's relevant conditions enable the general's order to cause the murderer's death via the causal process P_A, while Soldier B's relevant conditions enable the general's order to cause the murderer's death via the causal process P_B. This is a case of "over-enabling," where different sets of enablers are present and independently enable a cause to produce its effect, thereby activating multiple causal processes between the cause and the effect.

E1 fails to yield correct causal judgments in cases of over-enabling. On this account, Soldier A's relevant conditions do not enable the general's order to cause the murderer's death, because if Soldier A had not been standing by so that her relevant conditions had been absent, the general's order would still

[10] A complete eToC diagram incorporates further information, including possible derailers and safeguards of a causal process, as well as causal activities performed by the objects involved in that process. See Cartwright et al. (2026) for further illustrations.

have caused the murderer's death, by virtue of being enabled by Soldier B's relevant conditions. By the same reasoning, Soldier B's relevant conditions also fail to count as enablers. Taken together, these implications entail that nothing actually enables the general's order to cause the murderer's death. This result is problematic in two respects. First, it conflicts with the well-grounded judgment that both soldiers' relevant conditions play an enabling role in this case. Second, it commits us to the metaphysically implausible view that an event can cause its effect in the absence of any enabler when in fact some enablers or others are required.

The problem lies in the fact that E1 defines enablers as difference-makers for whether a cause produces its effect, that is, for whether a causal relation obtains. However, in cases of over-enabling, the same causal relation is realized simultaneously through multiple independent causal processes, each activated by a distinct set of enablers. Neither set of enablers can make a difference to whether the cause produces the effect due to the presence of the other enablers. To yield correct causal judgments in these cases, a plausible account of enabling must not assume that the obtaining of a causal relation counterfactually depends on the presence of its relevant enablers.

Is there an alternative account of enabling that better accommodates cases of over-enabling? Since each set of enablers is intimately tied to a specific causal process that it activates, I propose that an enabler should instead be understood as a difference-maker for some actual causal process through which a cause produces its effect, rather than for the causal relation between them. That is,

> **(E2)** *b* enables *c* to cause *e* in a situation *S* iff there exists in *S* an actually activated causal process *p* such that had *b* not occurred, *c* would not have caused *e* via *p*.

Such a process-based counterfactual account of enabling can yield correct causal judgments in cases of over-enabling. Had Soldier A not been standing by so that her relevant conditions had been absent, the general's order would not have caused the murderer's death via P_A. Thus, Soldier A's relevant conditions are enablers for the general's order to cause the murderer's death. Similarly, had Soldier B not been standing by so that her relevant conditions had been absent, P_B would not have been activated for the general's order to cause the murderer's death. Thus, Soldier B's relevant conditions are enablers, too.

E2 plausibly captures what enablers essentially are both because it ensures correct causal judgments in current cases and also because it withstands scrutiny when applied to new cases. In addition to over-enabling, preemptive enabling is another form of redundant enabling. In a case of preemptive enabling, multiple sets of potential enablers are present, but one set of potential enablers takes priority in enabling a cause to produce its effect and thereby prevents the others from doing so. Consider:

> **Soldier and Sergeant**: A soldier and a sergeant stand by, waiting for the general's order to execute a murderer. The policy is that the lower-ranking officer carries out the execution. The general shouts, "Execute!" The soldier fires at the murderer, resulting in a fatal wound. As a result, the murderer dies.

Both the soldier's relevant conditions and the sergeant's relevant conditions are potential enablers for the general's order to cause the murderer's death. Had only the soldier stood by, her relevant conditions

would have enabled the general's order to cause the murderer's death via a causal process P_O, which involves a sequence of events such as the soldier's firing, her bullet following a particular trajectory, and its eventually hitting the murderer. Had only the sergeant stood by, her relevant conditions would have enabled the general's order to cause the murderer's death via a causal process P_E, which involves a distinct sequence of events such as the sergeant's firing, her bullet following a particular trajectory, and its eventually hitting the murderer.

In the actual case, however, the policy assigning execution to the lower-ranking officer gives priority to the soldier's relevant conditions in serving as enablers. Thus, although both of their relevant conditions are capable of playing an enabling role on their own, only the soldier's relevant conditions actually enable the general's order to cause the murder's death. That is, P_O is activated as the only actual causal process. The sergeant's relevant conditions remain idle, leaving P_E, the alternative causal process, unactivated. Figure 2.2 represents the causal structure for the situation. Dotted ellipses represent the fact that a set of events capable of enabling the cause to produce the effect fail to actually do so, and dotted lines represent potential causal processes that are not actually activated in the situation.

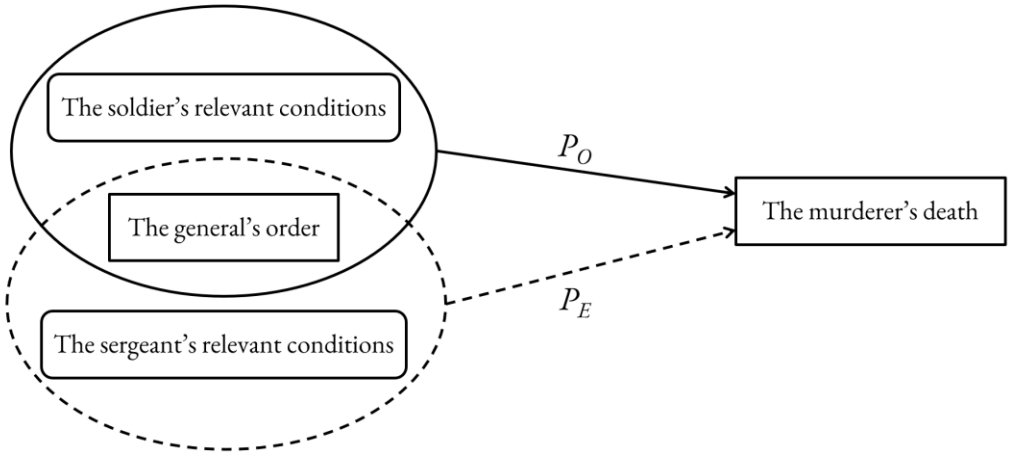


Figure 2.2 A case of preemptive enabling.

Again, E1 fails to yield correct causal judgments in cases of preemptive enabling. Had the soldier not been standing by so that her relevant conditions been absent, the general's order would have caused the murderer's death, by virtue of being enabled by the sergeant's relevant conditions. Thus, E1 falsely implies that the soldier's relevant conditions are not enablers in the actual situation. By contrast, E2 correctly implies that the soldier's relevant conditions are enablers in the situation described. For if the soldier's relevant conditions had been absent, the general's order would not have caused the murderer's death via the actual causal process P_O, even though it would have done so via an alternative causal process P_E. Also, E2 correctly implies that the sergeant's relevant conditions are not enablers in that situation. For if the sergeant's relevant conditions had been absent, the general's order would still have caused the murderer's death via the actual causal process P_O. No causal process activated in the actual situation would have become unactivated.

In sum, E2 is superior to E1 because, by construing enablers as difference-makers for causal processes underling causal relations, rather than for causal relations themselves, E2 can better handle cases in which the surface causal relation remains unchanged while its underlying causal process varies. This allows it to accurately identify redundant enablers, which actually play an enabling role even when alternative causal processes are counterfactually available, in both cases of over-enabling and those of preemptive enabling.

4. Facilitators as Redundant Enablers

Facilitators appear to play a causal role different from that of enablers. They accelerate or amplify a causal process, making the effect occur earlier or more severely. Unlike ordinary enablers, however, facilitators do not make an essential difference to whether a cause produces its effect at all. Does this show that facilitators and enablers are essentially distinct kinds of background conditions? No. As cases of redundant enabling indicate, enablers may likewise make no difference to whether a cause produces its effect when the cause would still have produced the effect via an alternative causal process in the relevant counterfactual situations. Therefore, it is unable to distinguish facilitators from enablers by consulting whether they make a difference to whether a causal relation obtains.

Indeed, facilitators are essentially redundant enablers. Suppose that $\mathbf{b} = \{b_1, b_2, ..., b_n\}$ is a set of enablers that enable c to cause e and that no facilitator for c to cause e is in $\mathbf{b}$. Let $P_{\neg F}$ denote the "unfacilitated" causal process by which c causes e, as activated by $\mathbf{b}$. Further suppose that f is a facilitator that, if present in the situation, will help c cause e more efficiently. My thesis is that, in cases where f is

present and facilitates *c* in causing *e*, its presence in fact activates a new causal process, P_F, through which *c* causes *e* such that, had *f* not been present, *c* would not have caused *e* via P_F. Thus, given E2, *f* enables *c* to cause *e*. Moreover, *f* is a redundant enabler because, had *f* not been absent, **b** would have enabled *c* to cause *e* via $P_{\neg F}$.

To illustrate, consider dish soap. I used dish soap when scrubbing a greasy plate. As a result, the grease was removed more easily than it otherwise would have. At the surface level, my use of dish soap, as a facilitator, appeared to accelerate "the" causal process through which my scrubbing caused grease removal. Nevertheless, a closer investigation reveals that its presence made a difference to the actual causal process in virtue of which this causal relation held. Had dish soap not been used, my use of a sponge and the presence of hot water would have enabled my scrubbing to cause grease removal via an "unfacilitated" causal process $P_{\neg DS}$, which consisted of key intermediate events such as:

(1) Hot water made the grease easier to detach from the plate surface by raising its temperature;

(2) My scrubbing with a sponge mechanically loosened the grease off the plate surface; and

(3) Water carried the loosened grease away.

In this case, grease removal would have depended entirely on heat, friction, and rinsing. For this reason, it would have taken a significant amount of time, water, and scrubbing efforts to remove the grease completely.

On the other hand, when dish soap was actually used, it significantly altered the causal process through which my scrubbing removed grease. Let P_{DS} denote the "facilitated" causal process in the presence of dish soap. Dish soap contains surfactant molecules with hydrophilic and hydrophobic parts, which can adsorb at the oil-water interface to reduce interfacial tension and help disperse grease into tiny droplets in water. Thus, compared to $P_{\neg DS}$, P_{DS} involves a significantly different intermediate event: the surfactant molecules in dish soap interacted with the grease and dispersed it into tiny droplets. These tiny droplets were then mechanically loosened from the plate surface by scrubbing and carried away by water. As a result, grease removal required less time, water, and scrubbing efforts than it would have in the absence of dish soap.

Figure 2.3 represents the relevant causal structure in the presence of dish soap. Clearly, had I not used dish soap, my scrubbing would not have caused grease removal via the actual causal process P_{DS}. Thus, given E2, my use of dish soap qualifies as an *enabler* that, together with other existing enablers such as my use of a sponge and the presence of hot water, enabled my scrubbing to cause grease removal. Moreover, it was a *redundant* enabler because had I not used dish soap, my use of a sponge and the presence of hot water would still have enabled my scrubbing to cause grease removal via a distinct causal process, $P_{\neg DS}$. The facilitating role of my use of dish soap is accounted for simply by the fact that, by proceeding via P_{DS}, my scrubbing exerted a faster and/or stronger causal influence on grease removal than it would have if it had proceeded via $P_{\neg DS}$. That is, P_{DS} has a greater rate of change and/or causal strength than $P_{\neg DS}$.

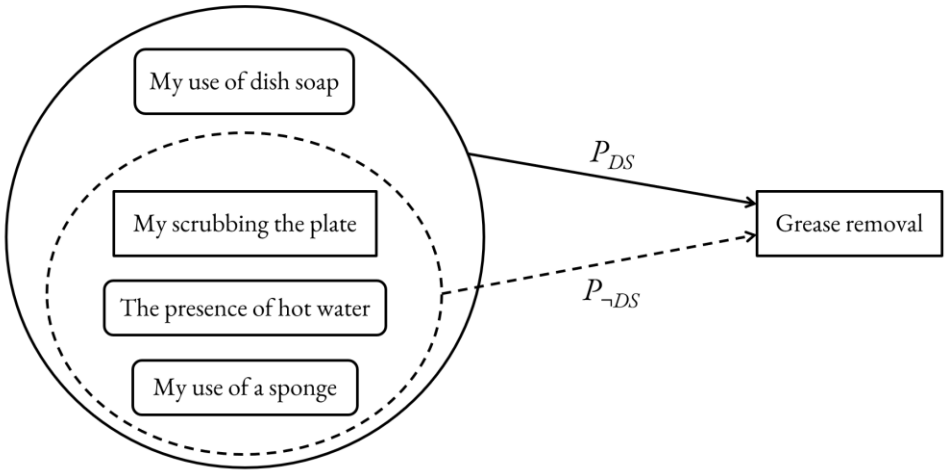


Figure 2.3 Dish soap as a redundant enabler.

Preemptive enabling and over-enabling are two major forms of redundant enabling. In parallel, a facilitator can display its facilitating role by serving as either a preemptive enabler or an over-enabler. My use of dish soap offers an instance of preemptive enablers that appear to display a facilitating role. In general,

> (**Facilitation by Preemptive Enabling**) *f* facilitates *c* to cause *e* in *S*, where other relevant enablers **b** = {b_1, b_2, ..., b_n} are present, by serving as a preemptive enabler iff there are two distinct causal processes, P_F and $P_{\neg F}$, available in *S* such that in *S*
>
> (i) **The Enabling Condition:** *f* and **b** (or a subset of **b**) jointly enable *c* to cause *e* via a causal process P_F;

(ii) **The Redundancy Condition:** Had f not been present, **b** would have enabled c to cause e via a distinct causal process $P_{\neg F}$; and

(iii) **The Facilitation Condition:** P_F has a greater rate of change and/or causal strength than $P_{\neg F}$.[11]

Figure 2.4 represents the causal structure of facilitation by preemptive enabling. Conditions (i) and (ii) jointly imply that f is a preemptive enabler. I assume a strict criterion for individuating causal processes: two causal processes P and P^* are identical iff they involve exactly the same set of intermediate events with the same time spacings arranged in the same sequence. Given this criterion, if f plays an enabling role in the activation of P_F, its presence must give rise to some intermediate event unique to P_F, which thereby makes P_F a distinct causal process from $P_{\neg F}$. In this case, although both P_F and $P_{\neg F}$ are available to c given the presence of the enablers required to activate each of them, c proceeds to cause e via P_F rather than $P_{\neg F}$. Presumably, this is because f is more causally active, such that it interacts with c, together with the other existing enablers, more quickly than those enablers would interact with c in the absence of f. Finally, condition (iii) ensures that e occurs earlier or with greater magnitude than it otherwise would if f had not been present, for which reason f can be viewed as a facilitator rather than, say, a decelerator or diminisher.

[11] Condition (i) can be further loosen by allowing f to work not with **b**, but with other events that are present in the actual scenario.

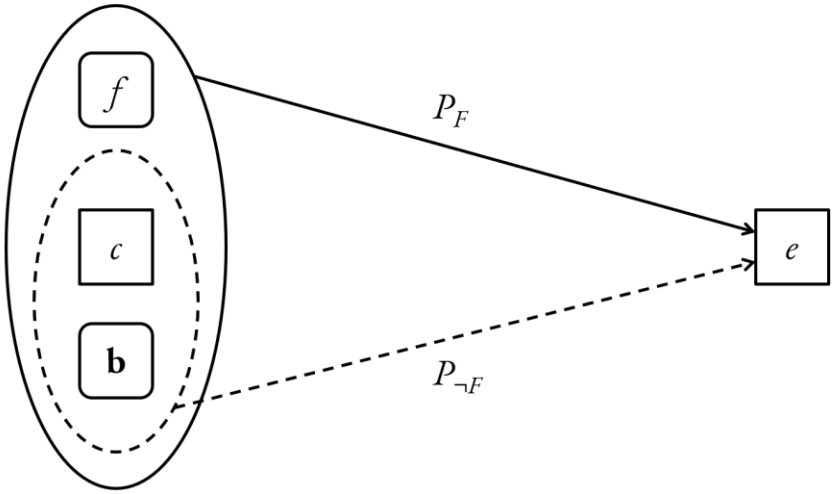


Figure 2.4 Facilitation by preemptive enabling.

Facilitation by preemptive enabling explains how most facilitators work, including catalysts in chemical reactions. Many chemical reactions can proceed through different reaction pathways, each consisting in a step-by-step sequence of intermediate chemical events through which reactants are transformed into products. Each reaction pathway has its own activation energy (i.e., the minimum energy needed for the reaction to proceed along it) and can operate only when the surrounding thermodynamic conditions are sufficient to overcome that energy barrier. Different reaction pathways for the same chemical reaction may thus be understood as alternative causal processes through which the presence of the reactants causes the formation of the products. When a catalyst is introduced to accelerate a chemical reaction, what it does is to provide a new reaction pathway from reactants to products that requires a lower activation energy. As a result, the reactants proceed preferentially along

this pathway rather than along other thermodynamically available pathways that require a higher activation energy. In this way, a catalyst functions as a preemptive enabler by making a difference to the actual causal process through which the reaction proceeds, even though the existing conditions would, in its absence, still have enabled the reaction to proceed via a distinct causal process.

For example, hydrogen peroxide (H_2O_2) can decompose into water and oxygen even without a catalyst, but in that case it proceeds through a reaction pathway that consists of a radical chain reaction. This pathway proceeds relatively slowly because it has a high activation energy. In the presence of manganese dioxide (MnO_2), by contrast, the decomposition proceeds mainly through a distinct surface-mediated reaction pathway, which has a much lower activation energy. In doing so, MnO_2 functions as a preemptive enabler. In its absence, the existing conditions would still have enabled H_2O_2 to decompose into water and oxygen via a distinct causal process. But its presence enables the reaction to proceed via a causal process with a much greater rate of change.

Besides serving as a preemptive enabler, a facilitator may play a facilitating role by serving as an over-enabler:

> (**Facilitation by Over-Enabling**) *f* facilitates *c* to cause *e* in situation *S*, where other relevant enablers **b** = {b_1, b_2, ..., b_n} are present, by serving as an over-enabler iff there are two distinct causal processes, P'_F and $P'_{\neg F}$, available in *S* such that in *S*
>
> (i) **The Enabling Condition:** *f* and **b** (or a subset of **b**) jointly enable *c* to cause *e* via a causal process P'_F;

(ii) **The Redundancy Condition: b** alone further enable *c* to cause *e* via a distinct causal process $P'_{\neg F}$; and

(iii) **The Facilitation Condition:** The causal strengths of *c* on *e* via P'_{F} and via $P'_{\neg F}$ are cumulative (though not necessarily additive) such that the total causal strength of *c* on *e* in the presence of *f* is greater than the causal strength of *c* on *e* in its absence, namely, via $P'_{\neg F}$ alone.

Figure 2.5 represents the causal structure of facilitation by over-enabling. Conditions (i) and (ii) jointly imply that *f* is an over-enabler in the sense that its presence activates a new causal process without interfering with the original one activated by **b**. Had *f* been absent, *c* would not have caused *e* via one of the actual causal processes, P'_{F}, but it would still have caused *e* via the other actual causal processes, $P'_{\neg F}$. Condition (iii) specifies that the new causal process activated by *f* can reinforce the causal strength of *c* on *e*. As a result, the presence of *f* increases the magnitude of *e* as produced by *c*. Here I restrict facilitation by over-enabling to causal strength, rather than rate of change, because only total causal strength can plausibly be understood as cumulative across the causal strengths of different causal processes.

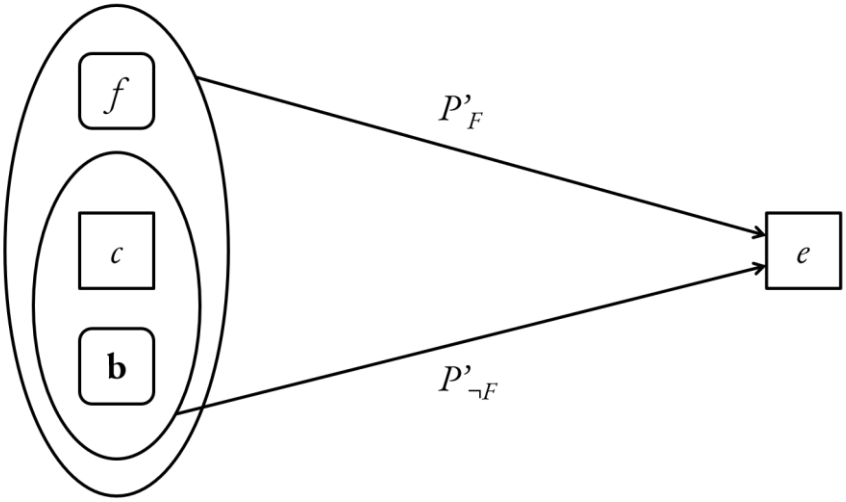


Figure 2.5 Facilitation by over-enabling.

For illustration, consider the causal relation between adrenaline administration and an increase in heart rate. Under normal circumstances, adrenaline raises heart rate through a primary causal process: it stimulates β1 receptors in the heart, enhancing myocardial contractility and conduction, thereby increasing heart rate. Now suppose that the patient has hyperthyroidism, a physiological condition in which adrenergic responses, including β2-mediated vascular effects, may be heightened. In this case, hyperthyroidism does not replace the original causal process, as adrenaline still increases heart rate through the β1-mediated pathway. However, in the presence of hyperthyroidism, adrenaline also opens up an additional causal process: it stimulates β2 receptors in peripheral blood vessels, causing vasodilation and a relative drop in blood pressure, which then triggers a baroreflex response that further increases heart rate. Thus, hyperthyroidism facilitates adrenaline administration in causing an increase

in the patient's heart rate by serving as an over-enabler. Had she not undergone hyperthyroidism, adrenaline administration might have increased her heart rate by, say, 20 bpm via the β1-mediated pathway. In the presence of hyperthyroidism, however, adrenaline still operates through the original β1-mediated pathway but also triggers an additional β2-mediated pathway, resulting in a larger overall increase in heart rate, e.g., by 40 bpm.

5. Background Conditions Partially Unified

The reduction of facilitators to (redundant) enablers not only illuminates the nature of facilitators but also has serious implications for a range of related issues. One issue it directly bears on is the extent to which background conditions are unified (or, disunified). Given that background conditions play a crucial role in determining whether and how a cause actually influences its effect, a good understanding of how many kinds of background conditions there fundamentally are is vital. On one extreme, it is conceptually possible to hold that:

> (**Monism about Background Conditions**) Fundamentally, there is only one kind of background condition: enablers.

This view is plainly implausible because there are well-accepted background conditions other than enablers. Preventors (or derailers) are background conditions that can block a cause from exercising its causal influence on a target effect, either by preventing it from triggering a causal process or by stopping a causal process already in train (Cartwright, 2021; Skow, 2023a). For instance, sudden rain is a preventor

that stops the lightning, which just struck a tree, from causing a forest fire, because had the rain not occurred, the lightning would have caused a forest fire.

Granted a pluralist view of background conditions, it remains an open question how plural they fundamentally are. Some may adopt a moderate pluralist view, holding that there are only a few distinct kinds of background condition, such as:

> (**Dualism about Background Conditions**) Fundamentally, there are two kinds of background condition: enablers and preventors.

Nevertheless, this view also seems doubtful because of background conditions that appear to play other causal roles. Facilitators, decelerators, and diminishers, for instance, appear to adjust merely a certain causal feature without making a difference to whether a cause produces its effect like typical enablers or preventors do. Indeed, for any causal feature p that characterizes how a cause influences its effect, it is possible to have background conditions that adjust only p and make no difference to whether the cause relation holds. The apparent diversity of such causal roles therefore supports a strong pluralist view of background conditions:

> (**Strong Pluralism about Background Conditions**) There are numerous kinds of background condition, each fundamentally irreducible to the others.

Strong pluralists might hold that, to the extent that there are a variety of causal features, there remains a diversity of background conditions beyond enablers and preventors.

However, if my argument that facilitators are redundant enablers is correct, we need not regard background conditions as too disunified. Rather, feature-adjusting background conditions can be understood as redundant enablers: they alter the actual causal process underlying a causal relation in a way analogous to facilitators. A redundant enabler may appear merely to adjust some causal feature associated with a causal relation, without affecting whether that relation obtains. But this is because it introduces a causal process that differs, with respect to that feature, from the default process that would have been activated by the other existing enablers in its absence. Thus, we can account for the nature of different kinds of feature-adjusting background conditions by replacing the facilitation condition in the previous accounts of facilitators with other suitable ones. For instance, a diminisher that appears to reduce the causal strength of *c* on *e* in the presence of other enablers **b** can be analyzed either as a preemptive enabler that replaces the default causal process activated by **b** with a new one that has a smaller causal strength, or as an over-enabler that activates an additional causal process with a negative causal strength without affecting the default process so that the total causal strength of *c* on *e* becomes smaller. A decelerator that appears to slow the rate at which *c* causes *e* in the presence of **b** can be analyzed as a preemptive enabler that replaces the default causal process activated by **b** with a new process operating at a slower rate of change. A similar analysis applies to other feature-adjusting background conditions as well. In this way, all feature-adjusting background conditions can be unified under enablers, thereby vindicating the dualist view.

Of course, denying that feature-adjusting background conditions are fundamentally distinct from ordinary enablers does not mean that we cannot retain a pragmatic distinction between them.

Non-redundant enablers, which make a difference to whether a causal relation actually obtains, may serve different predictive, explanatory, and manipulative purposes from those served by redundant enablers, which make a difference only to the actual causal process underlying that relation. This justifies our causal practice of treating them differently relative to our pragmatic goals. If the goal is to prevent a given outcome from occurring, one effective strategy is to remove possible non-redundant enablers that may be present in the situation. But if the goal is merely to mitigate the outcome, one may instead seek to introduce a redundant enabler that, in effect, diminishes causal strength.

6. Do Facilitators Cause?

Let's turn to another question: when *f* facilitates *c* in causing *e*, is *f* also a cause of *e* in the proper sense? The answer depends on what our best account of causation is. As we will see shortly, mainstream difference-making accounts of causation deny that facilitators are proper causes.

Many mainstream accounts of causation assume *the difference-making idea*, the view that causes make a difference to their effects (see also Lewis, 1973; Woodward, 2003; Menzies, 2004; Sartorio, 2005; Beebee et al. 2017; Andreas & Günther, 2021). However, they disagree about what kind of difference a cause needs to make. Here I consider three major difference-making accounts of causation:

(DM1) *c* causes *e* iff, had *c* not occurred, *e* would not have occurred.

(DM2) *c* causes *e* iff *c* is an INUS condition—i.e., an *i*nsufficient but *n*on-redundant component of an *u*nnecessary but *s*ufficient condition—for *e*.

(DM3) *c* causes *e* iff the occurrence of *c* raises the probability of the occurrence of *e*.[12]

All these accounts imply that, in a situation where my scrubbing would have removed the grease even without dish soap, my use of dish soap did not cause the grease removal in the proper sense, even though it did in fact help my scrubbing remove the grease more efficiently.

Start with DM1, the counterfactual account. Had I not used dish soap, the grease would still have been removed, due to my scrubbing as well as the presence of relevant enablers. Thus, according to DM1, my use of dish soap did not cause grease removal.

Next, consider DM2, the INUS account. The success of grease removal in the actual situation suggests that my scrubbing, my use of a sponge, the presence of hot water, and my use of dish soap jointly constituted a sufficient (though unnecessary) condition for grease removal. But my use of dish soap was not a non-redundant part of this condition, since the combination of the other three factors remained sufficient for grease removal. Thus, according to DM2, my use of dish soap did not cause grease removal.

Lastly, consider DM3, the probabilistic account. Did my use of dish soap raise the probability of grease removal *in the actual situation*? I take the relevant situation to hold fixed all actual events—including my scrubbing, my use of a sponge, and the presence of hot water—except my use of dish soap. Because the probability of grease removal in this situation was already 1 (that is, in the presence of all

[12] I ignore discussions of other prominent difference-making accounts of causation such as interventionism, because they will receive a treatment similar to the accounts I examine here.

actual events except my use of dish soap, the grease was bound to be removed), my use of dish soap could not further raise its probability. Thus, according to DM3, my use of dish soap did not cause grease removal.

It is thus fair to claim that mainstream difference-making accounts of causation do not regard facilitators as causes. If one accepts, as many do, the view that causes essentially make a difference to their effects, then one seems committed to denying that facilitators qualify as proper causes of the effects they facilitate.

This conclusion, if true, is more unwelcome than it first appears, for it contradicts several widely shared beliefs about causation. To start, because facilitators are, as I have argued, redundant enablers, they provide counterexamples to the claim that all enablers are proper causes. Since this claim is shared by both subjectivism and moderate objectivism about causal selection, both of which are widely supported, these views must either be rejected or significantly modified to accommodate facilitators. (In fact, the above line of reasoning applies to redundant enablers in general, not just to facilitators.)

Not merely so. What underlies both subjectivism and moderate objectivism, in my view, is a monist view about causal relevance:

> **(Monism about Causal Relevance)** For any two events *x* and *y*, *x* is causally relevant to *y* if and only if *x* causes (or prevents) *y*.

The guiding idea is that causation, i.e., the kind of relation between causes and effects, is the only fundamental kind of causal relevance. When *x* is causally relevant to *y*, *x*'s causal contribution to *y* can trace ultimately back to its being a cause of *y*. Thus, for any causal role *x* might play relative to *y* by serving as a background condition, this role is either a re-description of or derivable from *x*'s role as a cause of *y*.

Monism about causal relevance is appealing in virtue of the simplicity it ascribes to the fundamental causal structure of the world. If monism is true, then the causal structure of the world, in which innumerable causal factors are causally relevant to one another in arbitrarily complex ways, can be understood as constituted simply by a vast array of causes and effects. This simplified picture may help explain the widespread neglect of background conditions both in science, where causal inquiry focuses primarily on identifying the causes for a given effect and sidelines their relevant background conditions, and in philosophy, where causal relations are often represented simply as binary relations between a cause and an effect. On the monist view, such neglect may appear justified, since the causal contribution of background conditions can be ultimately reduced to their direct causal influence on the relevant effects qua causes. In fact, monism about causal relevance is so widely held that many philosophers who allege to offer an account of causal relevance are actually offering an account of causation (e.g., Woodward, 2003; Strand & Oftedal, 2019).[13]

---

[13] Those who adopt a mechanistic view of causation often distinguish between etiological relevance and constitutive relevance (Salmon, 1984; Craver, 2007; Craver et al., 2021). What is etiologically relevant to an event consists in its antecedent causes, whereas what is constitutively relevant to an event consists in the components (e.g., entities and activities) of the mechanism that gives rise to that event. For example, touching a worm's head is an antecedent cause of its backing up and is therefore etiologically relevant to the worm's behavior. By contrast, what is constitutively relevant to the worm's backing up is the components of the biological mechanism that

However, if, as facilitators suggest, not all enablers are causes, then causing and enabling should be seen as two fundamentally distinct kinds of causal relevance. That is, an event may enable another event to cause an effect without itself causing that effect, and vice versa. On this basis, monism about causal relevance is plainly false. We should instead adopt a pluralist view on which causing, enabling, and perhaps other kinds of causal relevance, if there are any, are equally fundamental.

7. Rescuing Monism about Causal Relevance

One simple way for monists about causal relevance to secure their view is to reject the difference-making idea. Although facilitators do not make a difference to the facilitated effects in a way specified by mainstream difference-making accounts of causation, they may still qualify as proper causes of those effects under some alternative account of causation that does not construe causes as difference-makers. Nevertheless, this approach is undesirable for those who support both monism about causal relevance and the difference-making idea. As I will argue in this section, the monist need not reject the difference-making idea to defend their view. In what follows, I consider four possible strategies by which a monist who embraces the difference-making idea might argue that facilitators are causes. The best strategy is to incorporate the notion of causal process, which is typical for mechanistic accounts of causation, with the difference-making idea. This results in a hybrid account of causation, which characterizes causes as

controls the movement of its muscles. If causal relevance is understood broadly to include not only etiological relevance but also constitutive relevance, then mechanists may not endorse monism about causal relevance. However, if causal relevance is understood more narrowly as encompassing only etiological relevance, then the mechanistic view of causation remains compatible with monism about causal relevance.

making a difference not to the occurrence of their effects, but to the actual causal processes leading to their effects.

1. *Argue by analogy*. To argue that facilitators cause the facilitated effects, the monist may draw an analogy with some event that merely hastens an effect but is nevertheless regarded as its cause. Suppose that Smith, unaware of her heart disease, planned to run in a marathon on Saturday, which would be fatal if she were to carry it out. Unfortunately, she died of a heart attack on Wednesday, brought on by an argument with her boss.[14] The argument hastened Smith's death without making a difference to whether or not it occurred, Nevertheless, we have a strong causal intuition that the argument caused her death. Insofar as this intuition is reliable, the monist may claim that we should similarly qualify my use of dish soap, a facilitator that hastened grease removal without making a difference to whether or not it occurred, as one of its causes.

This response ignores a crucial difference between the causal structure of Smith's death (Figure 2.6) and that of grease removal. Given Smith's heart disease, both her argument with her boss and her running a marathon could have caused her death if the other event had not occurred. In the actual situation, the argument occurred on Wednesday and activated a causal process leading to Smith's death, P_1. This thereby prevents the alternative causal process P_2, which would have been activated by her running a marathon on Saturday, from coming into existence. The dotted box indicates that the event

[14] I borrow this example from Mackie (1992). See also Bennett (1988), Lombard (1990), Sartorio (2006), and Touborg (2018) for the relevant discussion.

that Smith runs a marathon does not occur. We can verify whether the argument is an actual cause of Smith's death by holding fixed the non-occurrence of Smith's running a marathon as not occurring. In this case, whether the argument occurs makes a difference to whether Smith dies, according to all three difference-making accounts. (I leave the reasoning as an exercise for the reader.) Thus, the argument was a *cause* that hastens Smith's death.

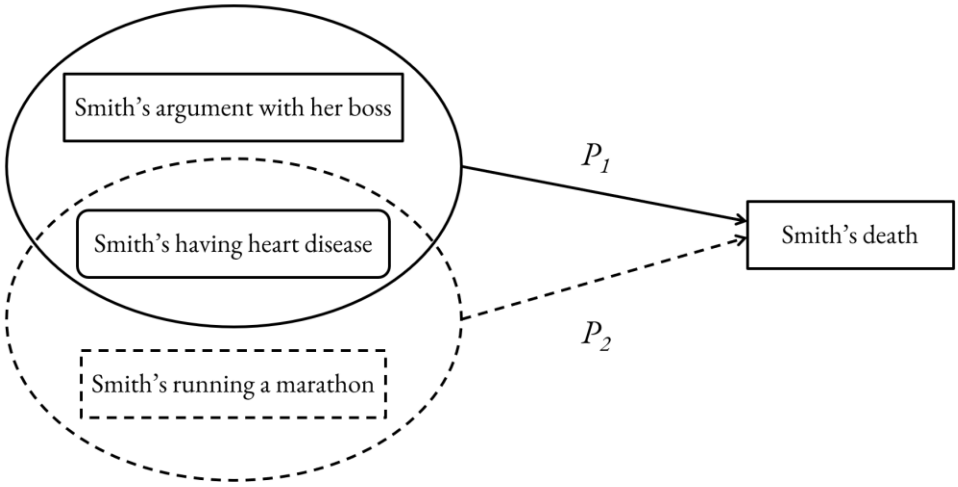


Figure 2.6. The causal structure of Smith's death.

By contrast, in the case of grease removal, my use of dish soap cannot be recognized as a difference-maker to whether the grease is removed on similar grounds. As Figure 2.3 suggests, my use of dish soap could activate a causal process leading to grease removal, P_{DS}, *only if* other relevant enablers are also present. Moreover, those enablers alone could have enabled my scrubbing to cause grease removal via a

distinct causal process, $P_{\neg DS}$. Thus, one cannot properly characterize my use of dish soap as a difference-maker to grease removal by holding fixed the situation in such a way that P_{DS} could be activated while the conditions needed to activate $P_{\neg DS}$ are absent. For this reason, my use of dish soap does not qualify as a proper cause in the way Smith's argument with her boss does.

2. *Make events fine-grained*. My analysis has assumed a relatively coarse-grained conception of event individuation, on which the time and magnitude of an event are merely accidental properties. This assumption allows us to treat the removal of grease in the actual situation, in which I used dish soap, and the removal of grease in the counterfactual situation, in which I did not, as one and the same event. On this basis, we can judge that my use of dish soap made no difference to whether that event occurred. This assumption, however, is not universally accepted. Some philosophers, for instance, take time to be an essential feature of events (Kim, 1976; Lombard, 1990). Similarly, scientists often characterize variables in a time-indexed manner, thereby representing events in a way that is sensitive to when they occur. Fewer philosophers and scientists adopt a comparable view about magnitude, but such a view is not inconceivable.

Given these considerations, the monist might insist that events are modally fragile, such that a subtle change in the timing or magnitude of an event results in a numerically distinct event. Since my use of dish soap affected when the grease was removed, it made a difference to whether the actual fine-grained event of grease removal occurred. It could therefore count as a cause of the actual event of grease removal, according to mainstream difference-making accounts of causation.

Nevertheless, whether, and to what extent, events are modally fragile remains highly controversial. For this reason, appealing to the fragility of events is not the most promising strategy for monists to defend their view. Even if coarse-grained individuation conditions for events are not universally accepted, always individuating events in a fine-grained way seems equally undesirable. As is often noted, if any minute difference makes an event numerically distinct, then an event will have too many causes, since many factors can easily make a difference to its occurrence under such fine-grained individuation conditions (Lewis, 1986c; Schaffer, 2005). For example, if the time at which the grease is removed is essential to the event, such that the event of the grease being removed at 13:00 is distinct from the event of its being removed at 12:59, then my getting up five minutes late this morning, my eating lunch in the living room rather than the dining room, the dish soap's being placed less than one meter away from me, and many other apparently irrelevant events would all count as causes of the actual event of grease removal. The counter-intuitiveness of this result suggests that whether an event should be individuated in a fine-grained way may be, at best, a contextual matter. At least for those who think that our metaphysical commitments should align with our robust causal intuitions, events should not be treated as modally fragile in all contexts. In contexts where coarse-grained individuation of events is more appropriate, facilitators do not qualify as genuine causes, because they make no difference to whether the effects they facilitate occur.

3. *Contrastivize effects*. Mainstream difference-making accounts assume that causation is a binary relation. Under this assumption, it is natural to interpret the difference-making idea in terms of the occurrence of causal relata: *c* causes *e* iff *c*'s occurrence makes a difference to *e*'s occurrence. Mainstream

difference-making accounts differ in how they specify the relevant kind of difference. However, the monist may argue that this occurrence-based conception of difference-making is too narrow, since it conceptually excludes the possibility that a cause makes a difference to its effect in ways other than by making a difference to whether the effect occurs. Moreover, this conception fails to accommodate the contrastive character of many causal explanations, in which an actual outcome is explicitly contrasted with some alternative outcome and the explanandum is why the former occurred rather than the latter (e.g., why did World War I break out in Europe rather than Asia?). For these reasons, the monist may instead adopt a contrastive account of causation, according to which causation is a quaternary relation: the contrast between an actual cause *c* and some contrast event *c*' causes the contrast between an actual event *e* and some contrast event *e*':

> (**DM4**) *c* rather than *c*' causes *e* rather than *e*' iff, had *c*' rather than *c* occurred, *e*' rather than *e* would have occurred (Schaffer, 2005; Northcott, 2008).[15]

On this account, causal relations expressible in the form "*c* causes *e*" are merely a specific type of contrastive causation: the occurrence of *c* rather than its non-occurrence causes the occurrence of *e* rather than its non-occurrence. The monist may claim, in the case of facilitators, the relevant causal relation belongs to a different type: the presence of a facilitator rather than its absence causes the effect to occur at an earlier time *t* rather than at a later time *t*', or to have magnitude *m* rather than a smaller magnitude

[15] I am talking primarily about a contrastive account of causal relation, not of causal claim or of causal explanation.

m'. For example, in the dish soap example, the relevant causal relation is that my use of dish soap rather than not using it caused the grease to be removed earlier rather than later. By properly specifying a contrast event associated with the actual effect, the monist can identify my use of dish soap, or any other facilitator, as a cause of the effect it facilitated.

Although this response allows facilitators to be identified as causes, I suspect that adopting a contrastive account of causation is too high a price to pay. I will focus on one of its difficulties. DM4 implies that an actual event *c* and an unactualized event *c*' can form a proper contrast in a causal relation only if there is a corresponding contrast between an actual event *e* and an unactualized contrast event *e*'. If no proper *e*-*e*' pair can be identified, then *c* cannot be a cause of *e* relative to *c*'. However, this implication is false. Suppose that, at 10 am, the king drank a cup of tea poisoned by Assassin 1 and died immediately. Had Assassin 1 failed to poison the tea, Assassin 2 would have poisoned it with exactly the same kind of poison. The king would therefore still have died at the same time and in exactly the same way. It is perfectly natural to say that Assassin 1's poisoning, rather than Assassin 2's poisoning, caused the king's death. However, since the two poisonings would have produced the very same outcome, there is no suitable contrast event on the effect side to form a causal relation in the way required by DM4. The contrastive account of causation therefore fails to provide a satisfactory difference-making account that can identify Assassin 1's poisoning, in contrast to Assassin 2's poisoning, as a cause of the king's death. This may explain why many philosophers who adopt a contrastive approach apply it only to causal explanation rather than also to causation itself (van Fraassen, 1980; Lipton, 1990; Hitchcock, 1999;

Sober, 2020). For the claim that causal explanations are contrastive is compatible with the view that causal relations are not essentially contrastive.

4. *Reanalyze causes*. Ideally, a satisfactory defense for monism about causal relevance should be able to (a) identify facilitators as causes, (b) represent causation as a binary relation, and (c) preserve the widely shared difference-making idea of causation. To satisfy (a), the monist must reject mainstream difference-making accounts of causation, which characterize a cause as making a difference to whether or not its effect occurs. To satisfy (b), the monist must not appeal to accounts that characterize causation as essentially contrastive, that is, accounts that analyze the difference a cause makes to its effect in terms of a contrast between the actual effect and an alternative event. To satisfy (c), however, the monist must still embrace a difference-making account of causation, which characterizes a cause as making some kind of difference to its effect. Can the monist satisfy (a)–(c) simultaneously?

She can. Christian Loew (2019) recently articulates and defends an alternative difference-making account of causation, which characterizes a cause as essentially making a difference to some actual causal process that leads to its effect:

> (**DM5**) *c* causes *e* iff, had *c* not occurred, some actual causal process leading to *e* would not have occurred.

This account, in my view, is more plausible than mainstream difference-making accounts of causation because it is more inclusive in properly identifying events as causes. We have seen cases of preemptive causation (e.g., both the case of Smith's death and that of the king's death), where it is intuitively

appealing to regard an event as a cause of an outcome when it helped bring into existence the particular causal process leading to that outcome, even though it made no difference to whether that outcome occurred. Compared with mainstream difference-making accounts, which interprets difference-making in terms of the occurrence of causal relata, DM5 can correctly identify Smith's argument with her boss as a cause of her death and Assassin 1's poisoning as a cause of the king's death.[16]

Given its independent plausibility, the monist can appeal to DM5 to assess the causal status of facilitators. Although facilitators, as redundant enablers, make no difference to whether the coarse-grained effects they facilitate occur, they do make a difference to the actual causal processes through which those effects are produced. Had I not used dish soap, the actual causal process leading to grease removal would not have occurred. Instead, a different causal process would have produced grease removal at a later time. DM5 therefore identifies my use of dish soap as a cause of grease removal. The same analysis applies to other facilitators. In this way, the monist has a response that satisfies (a)–(c). This suggests that, beyond DM5's independent plausibility, those who accept monism about causal relevance have a further reason to prefer DM5 over mainstream difference-making accounts of causation.

8. Conclusion

I have motivated an account of enabling as well as an account of causation. To accommodate cases of redundant enabling, enablers should be understood as difference-makers for the actual causal

[16] See Loew (2019) for plausible cases along this line.

processes through which causes produce their effects. To defend monism about causal relevance while preserving the difference-making idea, causes should be understood as difference-makers for the actual causal processes through which effects are produced. By taking difference-making with respect to causal processes as a shared defining feature, the pair of accounts ontologically unify causation and enabling. If *b* enables *c* to cause *e*, then, given that *b* makes a difference to some actual causal process by which *e* is produced, *b* is also a cause of *e*. Conversely, if *b* causes *e*, then, insofar as some event enables *b* to cause *e*, *b* also enables that event to cause *e*. Hence, the class of enablers is identical to the class of proper causes.

One might object that the two accounts are overly inclusive, since many events will count as the enablers and proper causes of a given effect so long as they make some difference, however slight, to the actual causal process by which that effect is produced, in however tiny way. I take this inclusiveness to be a virtue rather than a defect. The accounts provide a broad yet firm boundary for events that are causally relevant to an effect. This is compatible with more stringent accounts that distinguish explanatorily significant causes and enablers from other proper causes and enablers. For example, one might hold that only causes that make a difference to the occurrence of an outcome are explanatorily relevant. However, it is beyond the scope of this chapter to provide such accounts.

One nice consequence of this pair of accounts is that they help us better understand the causal status of facilitators, as well as that of many other background conditions that appear to differ from ordinary enablers. On the one hand, facilitators are essentially redundant enablers, which activate an alternative causal process that either substitutes for or supplements the default causal process through which a cause produces its effect. The two distinct causal structures, facilitation by preemptive enabling

and facilitation by over-enabling, each involve a distinctive facilitation condition whose satisfaction gives rise to the apparent facilitating role of facilitators. On the other hand, facilitators cause the facilitated effects because they make a difference to the actual causal processes leading to those effects. In a causal explanation, the explanatory purpose determines whether facilitators, as causes, are explanatorily relevant. When one is interested in why an effect occurs, facilitators are likely to be explanatorily irrelevant compared with causes that make a difference to whether the effect occurs. But when the explanandum concerns the timing or magnitude of an effect, facilitators may be foregrounded as important causes.

The analysis for the causal status of facilitators can be extended to other apparently distinct kinds of background conditions. Like facilitators, these background conditions function both as redundant enablers and as proper causes insofar as they shape the actual causal process that leads to an effect. Once we take into account the full range of background conditions that are reducible to redundant enablers, the thesis that enablers are proper causes turns out to be much more far-reaching and better justified than it initially appears.

Chapter 3

Causal-Structural Explanation

Co-authored with Zili Dong

1. Introduction

We often distinguish causal explanations from non-causal ones. Compare the two explanations:

(1) John has lung cancer because he smoked a lot.

(2) The mother cannot evenly distribute twenty-three strawberries among her three children without cutting any because twenty-three cannot be evenly divided by three (Lange 2013).

One can easily discern that (1) is a causal explanation while (2) is not. But on what grounds are causal explanations distinguished from non-causal ones?

To better frame this question, it is helpful to think of an explanation as consisting of three components: the *explanandum* to be explained, the *explanans* that provides explanatory information, and the *dependence relation* between them that captures how the two are connected (Ross, 2025). Which component determines whether an explanation is causal or not? Two options emerge naturally:

(The Explanans View) An explanation is causal if and only if the explanans is causal (that is, the explanans provides causal information to account for the explanandum).

(The Relation View) An explanation is causal if and only if the dependence relation between the explanatory relata is causal (that is, the explanandum causally depends on the explanans).

To our knowledge, this distinction has not been clearly made in the literature. The received conception of causal explanation in philosophy of science and metaphysics states that to causally explain a phenomenon is to provide information about its relevant causes (Lewis, 1986b; Woodward, 2003; Lipton, 2009; Pincock, 2018; Skow, 2023b; Ross, 2025). David Lewis, for instance, famously claims that "to explain an event is to provide some information about its causal history" (Lewis, 1986b: 217), where the causal history of an event mainly includes the events it causally depends on, i.e., its causes. Bradford Skow makes it more explicit that "a causal explanation is an answer to 'Why X?' that says something about the causes of X" (Skow, 2023b). For those who accept this received conception, the explanans view and the relation view collapse into one: the explanans can causally inform us why the explanandum occurs just in case it depicts the causes of the explanandum, which amounts to saying that the explanandum causally depends on the explanans.

The main goal of this chapter is to reject the received conception of causal explanation and substantiate the debate between the explanans view and the relation view. To demonstrate a genuine disagreement between the two views, we identify a distinctive kind of explanation, *causal-structural*

explanation, that has been frequently employed in scientific practices of causal inference but does not appear to align well with the received conception of causal explanation. After analyzing the explanatory pattern of causal-structural explanations, we argue that while the explanans view will classify these explanations as causal explanations, nevertheless, the relation view will classify them as non-causal explanations. There is no inconsistency here, as it makes perfect sense for a causal-structural explanation to invoke both causal and non-causal features in achieving its explanatory success.

In a causal-structural explanation, the phenomenon to be explained is a set of observed correlations among events or variables (hereafter "causal factors" for simplicity).[17] These correlations are accounted for by appealing to a causal structure whose structural pattern (and, where applicable, those causal relations that instantiate the pattern) is posited to generate the observed correlations. In addition, a causal-structural explanation must assume certain constraining principles that govern how the appealed causal structure can generate the correlations to be explained, even though these constraints often remain implicit in an explanation.[18] Strikingly, as we shall argue, the dependence relation in a causal-structural explanation, as captured by the constraining principles related to an explanatory causal structure, is *modally stronger* than mere causal dependence. Thus, causal-structural explanations exhibit

[17] To simplify the discussion, we assume that the explanans in a causal-structural explanation is population-level probabilistic correlations. This way, we rule out those associations due to sampling error or small sample size as interesting or relevant explananda.

[18] The concept of constraint used here can be understood along the line of Green and Jones (2016) (cf. Lange 2023). According to Green and Jones (2016: 346), such constraints operate at a high level of abstraction which "make some causal details irrelevant for understanding why the systems exhibit a general pattern."

a *dual* nature: while the explanans causally informs why the correlations to be explained obtain, these correlations depend non-causally on the explanans.

We begin by illustrating the notion of causal-structural explanation with an example from epidemiology (Section 2). We then analyze the explanatory relata of causal-structural explanations (Section 3) and the dependence relation between them (Section 4). Finally, we highlight the distinctive dual nature of causal-structural explanations (Section 5).

2. The Birthweight Paradox Explained

Maternal smoking is a known risk factor for both low birthweight and infant mortality. However, when low-birthweight infants born to smoking mothers are compared with those born to non-smoking mothers, a puzzling pattern emerges: the former group exhibits a lower infant mortality rate than the latter. This phenomenon is known as the "birthweight paradox," as the correlation between maternal smoking and infant mortality appears to reverse unexpectedly when conditioned on low birthweight (see Figure 2 in Hernández-Díaz et al., 2006).

Epidemiologists seek to resolve the birthweight paradox by identifying the causal structure that gives rise to, and thus explains, the following correlations between maternal smoking (*Smoking*) and infant mortality (*Mortality*):

(i) *Smoking* and *Mortality* are positively correlated at the marginal level.

(ii) *Smoking* and *Mortality* are positively correlated conditional on middle or high birthweight.

(iii) *Smoking* and *Mortality* are negatively correlated conditional on low birthweight.

One widely accepted explanation of these correlations appeals to a causal structure involving four variables: *Smoking*, *Mortality*, *LBW* (i.e., low birthweight), and *Defect* (i.e., birth defects) (Hernández-Díaz et al., 2006). This causal structure is represented as the graphical causal model in Figure 3.1, where each directed edge denotes a causal relation between two variables.

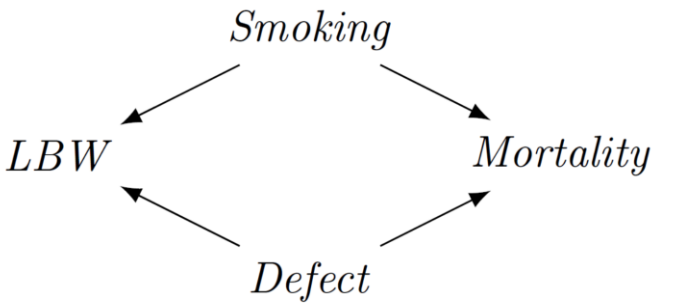


Figure 3.1 The causal structure underlying the birthweight paradox.

In this causal model, correlations (i) and (ii) are explained by the direct causal relation between *Smoking* and *Mortality*, together with the absence of causal relations strong enough to overturn these positive correlations. Correlation (iii), by contrast, is mainly explained by the fact that *LBW* is a *common effect* of *Smoking* and *Defect*. Birth defects are risk factors for both low birthweight and infant mortality.

Crucially, infants with birth defects have a higher mortality rate than those born to smoking mothers. By focusing on low-birthweight infants, we are in effect conditioning on the common effect *LBW*, thereby inducing a spurious negative correlation between *Smoking* and *Defect*. This is also known as "collider bias" in statistics—a collider is a variable that is affected by two or more variables (see Hernán & Monge, 2023). In our case, the spurious negative correlation between *Smoking* and *Defect* emerges because when an infant's low birthweight is not caused by maternal smoking, it is more likely to result from birth defects. Consequently, given that birth defects lead to a higher infant mortality rate than maternal smoking, *Smoking* and *Mortality* appear negatively correlated among low-birthweight infants.

The explanation of correlations (i)–(iii), which constitute the birthweight paradox, exemplifies what we call a "causal-structural explanation," in which a set of observed correlations are accounted for by the causal structure that generates them. The search for this kind of explanation is a key objective not just in epidemiology, but also in many other scientific domains such as physics (Gallus et al., 2023). Nevertheless, to our knowledge, no systematic analysis of causal-structural explanations has yet been offered. Although the significant role of causal structures in scientific explanation has been recognized in philosophy of science, mainstream philosophical discussions of causal explanation still treat the explanatory relata primarily as events or variables, thereby sidestepping scientific explanations that involve other kinds of explanatory relata. Moreover, recent advances in causal modelling focus primarily on developing methods that infer from statistical data to causal structures that putatively explain the data (e.g., Spirtes et al., 2000; Pearl, 2009). But these methodological developments do not by themselves suggest a philosophical account of the ontological and explanatory character of explanations that appeal

directly to causal structure. These considerations justify the explicit recognition of causal-structural explanation as an important kind of scientific explanation, as well as the need for a systematic philosophical treatment of its explanatory pattern.

3. The Explanatory Relata

In our view, a causal-structural explanation exhibits the following explanatory pattern:

(Causal-Structural Explanation) A causal structure *S* explains a set of correlations *C* if and only if:

(1) The explanandum identifies *C*'s specific features *F* that call for explanation;

(2) The explanans draws on *S*'s relevant features that give rise to *F*, including:

   a) *S*'s structural pattern *P*, and

   b) The causal relations *R* that constitute *S* (whether to include *R* or not will depend on which features of *C* are to be explained);

(3) *F* counterfactually depends on *P* and, where relevant, on *R*. This counterfactual dependence is regulated by a set of constraining assumptions *A*, which dictate how the explanandum features of the correlations depend on the explanans features of the causal structure (see Figure 3.2).

We will examine the explanatory relata of causal-structural explanations in this section and turn to the dependence relation between them in the next section.

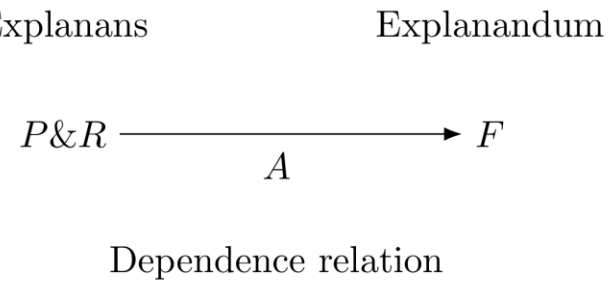


Figure 3.2 The explanatory pattern of causal-structural explanation: the explanandum (the features *F* of the correlations to be explained), the explanans (the structural pattern *P* of the explanatory causal structure and the causal relations *R* that constitute this causal structure), and the dependence relation (the assumptions *A* that constrain how *F* depend on *P* and *R*).

For ease of discussion, we focus primarily on causal structures that can be represented as structural causal models (SCMs), since this framework is arguably the most well-developed approach for representing and inferring causal structures and is widely employed across different scientific domains. In an SCM, a causal structure is depicted by a directed acyclic graph (DAG) that represents the causal relations among variables (e.g., Figure 3.1), along with a set of structural equations specifying how each effect (or, endogenous) variable is determined by its cause variables. However, our focus on causal structures representable as SCMs is primarily for illustrative purposes. We acknowledge that some causal structures (e.g., causal structures in fundamental physics or complex non-linear systems) may lie beyond

the representational capacity of SCMs. For example, even if Bell's (2004) local causality assumption is violated in the quantum realm, making SCMs unapplicable, we can still talk about the explanation of quantum correlations in terms of non-classical causal structures. We will not explore these special cases in detail, except to note that our proposed account of causal-structural explanation applies to these causal structures as well.

The explanandum in a causal-structural explanation consists, by definition, of one or more correlations among a set of causal factors. Such correlations may range from the co-occurrence of two events (e.g., the observed co-occurrence of a rising barometer and an approaching storm) to complex patterns of correlations among multiple variables (e.g., those observed in the birthweight paradox). Upon observing a set of correlations, one may ask "Why do they obtain?" at varying levels of granularity. A coarse-grained explanandum is concerned solely with the presence of certain probabilistic (in)dependence relations among causal factors of interest, without attending to their specific qualitative or quantitative features (e.g., direction, strength, robustness). A more fine-grained explanandum, by contrast, incorporates some of these features and seeks to explain them as well. For instance, in resolving the birthweight paradox, the explanandum includes not only whether *Smoking* and *Mortality* are correlated at the marginal or conditional level, but also the direction of each correlation between them.

The explanans in a causal-structural explanation comprises those features of a causal structure that account for the explanandum. These include both the *abstract* structural pattern of a causal structure and the *concrete* causal relations that instantiate this pattern. The two components play distinct explanatory roles in causal-structural explanations.

The structural pattern of a causal structure is the overall abstract pattern exhibited by the presence (or absence) of causal relations among its causal factors. This means that two distinct causal structures over different sets of causal variables can share the same structural pattern if the two sets of variables are arranged in isomorphic patterns of causal dependence. For instance, consider the causal structure that genetic mutation (*Mutation*) and alcohol use (*Alcohol*) independently cause lung cancer (*Cancer*), represented as $Mutation \rightarrow Cancer \leftarrow Alcohol$, and the causal structure that family socioeconomic status (*SES*) and intelligence (*Intelligence*) independently influence one's education level (*Education*), represented as $SES \rightarrow Education \leftarrow Intelligence$. Although the two causal structures involve different variables and causal relations, they share the same abstract structural pattern: $X \rightarrow Y \leftarrow Z$.

Different levels of granularity in the explanandum impose different explanatory demands on which features of a causal structure need to be cited in the explanans. Within the framework of SCMs, a coarse-grained explanandum targeting only the presence of certain probabilistic (in)dependence relations while ignoring their qualitative or quantitative details can be adequately explained by appealing solely to the structural pattern of a causal structure containing all relevant variables. This is because the structural pattern of a (causally sufficient) causal structure alone can determine whether any two of its variables are probabilistically independent or not, though it is insufficient to determine more fine-grained probabilistic information such as the direction or strength of a correlation.

Consider three basic structural patterns:

(a) Fork: $X \leftarrow Y \rightarrow Z$, representing Y as a common cause of X and Z;

(b) Chain: $X \rightarrow Y \rightarrow Z$, representing Y as a mediator connecting X with Z;

(c) Collider: $X \rightarrow Y \leftarrow Z$, representing Y as a common effect of X and Z.

Causal structures in real life are rarely this simple. However, understanding how these simple patterns generate probabilistic (in)dependence relations among the relevant variables provides an intuitive ground for understanding more complex structural patterns. In fact, it has been shown that there exists a criterion, d-separation, that allows us to decode the probabilistic behavior of a graphical causal model of any complexity. For the sake of space, we will not dive into the details here. [19]

Each of the three structural patterns generates a specific set of probabilistic (in)dependence relations among the relevant variables. Consider, for example, the probabilistic relation between X and Z. In both the fork structure and the chain structure, X and Z are marginally dependent but become independent when conditioned on Y. This is what it means when it is said that holding Y fixed can "explain away" or "screen off" the correlation between X and Z. In contrast, in the collider structure, X and Z are marginally independent but become dependent when conditioned on Y.

However, when the explanandum involves more fine-grained probabilistic information than the mere presence of probabilistic (in)dependence relations, such as the direction of correlation in correlations (i)–(iii), the structural pattern of a causal structure alone cannot provide an adequate explanation. In a causal-structural explanation for fine-grained probabilistic information about

[19] See Pearl (2000: §1.2.3) and Pearl, Glymour & Jewell (2016: §2.4) for a detailed discussion of d-separation.

correlations, the structural pattern of a causal structure and the concrete causal relations that constitute this causal structure play complementary explanatory roles: the former accounts for why certain correlations (or independences) can hold among variables, while the latter explains why these correlations display fine-grained probabilistic features such as their direction and strength. For instance, to explain why *Smoking* and *Mortality* are negatively correlated among low-birthweight infants, one must appeal not only to the information about the structural pattern of the causal structure depicted in Figure 3.1 (such as the fact that *LBW* and *Mortality* are common effects of *Smoking* and *Defect*), but also to the information specific to the causal relations within this causal structure, especially the respective causal strengths of *Smoking* and *Defect* on *Mortality*.

Moreover, concrete causal relations within a causal structure can offer evidential support for the plausibility of a causal-structural explanation. A single set of correlations can often be accommodated and thus explained by multiple candidate causal structures. When there are rival explanations for the same correlations, the structural pattern by itself hardly determines which causal-structural explanation we should accept. In this case, it is necessary to go beyond the structural pattern of a causal structure and examine the plausibility of individual causal relations it contains. For example, the observation that two variables, say P and Q, are marginally correlated but conditionally independent given a third variable O can be explained at least by three different causal structures: $P \leftarrow O \rightarrow Q$, $P \rightarrow O \rightarrow Q$, or $Q \rightarrow O \rightarrow P$. Choosing among these alternatives causal-structural explanations requires assessing the direction of the causal relation between P and O, as well as that between Q and O. When the causal relations posited by

one candidate causal structure are more plausible than those posited by its rivals, we have strong reason to prefer an explanation appealing to that causal structure over the others.

4. The Dependence Relation

In a causal-structural explanation, the explained correlations *counterfactually* depend on the explanatory causal structure: were the causal structure been different, the correlations would have been different. For example, if what we have among X, Y and Z is not a chain structure ($X \rightarrow Y \rightarrow Z$), but a collider structure ($X \rightarrow Y \leftarrow Z$), the correlations among these variables would have been different. Note that this is not a *causal* counterfactual since the causal structure does not cause the correlations, and it is also unclear that it can be understood in terms of Woodward's (2003) interventionist semantics. As we shall see in the next section, it is precisely because this counterfactual dependence is non-causal that we claim causal-structural explanation has a dual nature. Still, the question remains: how should we understand this dependence relation between the relevant features of a causal structure and the observed correlations among its variables? Regardless of the ontological nature of such dependence, what we can say is that it is constrained by certain robust principles or assumptions. These constraints endow the dependence with explanatory power in a way similar to how causal laws endow causal dependence with explanatory power. While often implicitly assumed in causal-structural explanations, these constraints must be explicitly acknowledged to ensure a complete causal-structural explanation.

For causal structures representable by SCMs, two important constraints they are subject to are the *causal Markov condition* (CMC) and the *causal faithfulness condition* (CFC).[20] Simply put, CMC states that for any variable X in a (causally sufficient) causal structure representable by a DAG, X is probabilistically independent of every other variable except X's (direct or indirect) effects, conditional on X's direct causes. CFC states that a causal structure permits only those conditional independence relations implied by CMC (Zhang & Spirtes, 2008). The conjunction of CMC and CFC has implications for what probabilistic (in)dependence relations that a causal structure can be expected to generate by virtue of its structural pattern. For example, CMC implies that in a fork structure $X \leftarrow Y \rightarrow Z$ or a chain structure $X \rightarrow Y \rightarrow Z$, X and Z are independent conditional on Y, while CFC implies that X and Z are marginally correlated. CMC and CFC together also imply that in the causal structure in Figure 3.1, *Smoking* and *Defect* are conditionally correlated given *LBW*, which plays a key role in the explanation of the birthweight paradox. For this reason, a complete explanation of this paradox must explicitly recognize the role CMC and CFC play in the generation of the relevant correlations.

It is worth asking further where constraints such as CMC and CFC come from. One may treat them merely as methodological assumptions posited in SCMs to enable the inference of causal structures from probabilistic data (Spirtes et al., 2000). By contrast, we view these constraints as substantive features exhibited by a wide range of real-world causal structures (see also Weinberger et al., forthcoming).

[20] Other important constraints in SCMs include minimality, modularity, sufficiency, and so on (see Spirtes et al., 1993; Pearl, 2009).

It is precisely by exhibiting such features that a causal structure becomes representable by a SCM. In acknowledging this, we admit the possibility that some causal structures—such as those in quantum physics or those involving opposite causal influences that cancel out each other—may not conform to CMC and/or CFC, and thus fall outside the representational scope of SCMs (Hesslow, 1976; Cartwright, 1999). These causal structures may instead be governed by alternative constraints that delimit the kinds of observable correlations they can give rise to. Consequently, when one appeals to a causal structure that cannot be represented by an SCM to explain certain correlations, a complete explanation should cite these alternative constraints to justify the dependence relation between the explanans features of the causal structure and the explanandum features of the correlations.[21]

5. A Dual Nature

Traditionally, the kind of information provided by the explanans is taken to align with the kind of dependence relation between the explanatory relata. In a typical causal explanation, the explanans provides causal information to account for the explanandum, by identifying factors that it causally depends on. By contrast, in a typical non-causal explanation, the explanans provides non-causal information to account for the explanandum by identifying factors that it non-causally (e.g., mathematically or ontologically) depends on. It is thus widely believed that a clear boundary exists

[21] See, for example, Costa and Shrapnel (2016) for alternative constraints that govern causal models for quantum systems.

between causal explanations and non-causal explanations. If this is true, there are no quarrels between the explanans view and the relation view.

However, the two views diverge on causal-structural explanations because of their *dual* nature. On the one hand, identifying a causal structure whose relevant features give rise to certain observed correlations provides *causal* information about why such correlations obtain. In this regard, causal-structural explanations resemble typical causal explanations in that the explanans causally informs why the explanandum occurs. Thus, if the explanans view is correct, causal-structural explanations should count as a kind of causal explanation.

On the other hand, causal-structural explanations also exhibit features characteristic of typical non-causal explanations because the dependence relation between the explanatory relata is modally stronger than causal. To see this, consider the causal structure *Mutation* → *Cancer* ← *Alcohol*. By virtue of its collider structural pattern and the constraints imposed by principles such as CMC and CFC, this causal structure generates, and thus explains, the correlation between *Mutation* and *Alcohol* among individuals with lung cancer as well as their independence in the overall population.

The dependence relation between the observed probabilistic relations between *Mutation* and *Alcohol* and the relevant features of the underlying causal structure cannot be fully captured in causal terms. One reason is that it goes against our linguistic and metaphysical intuitions to identify a causal structure as a *cause* of the correlations it generates. Our causal discourse admits events, variables, properties, etc. as causal relata. But it never admits causal structures as causal relata. Moreover, a causal

structure consists of a collection of causal relations, which makes it a category mistake to identify causal structures as causes.

More importantly, the relevant features of a causal structure determine the probabilistic relations among its variables to an extent that causes cannot achieve on their effects. It is widely accepted that two factors can be causally related, only if they are "fully distinct" or "independently manipulable" in the sense that one factor can, in principle, attain any of its possible states when the other factor is held fixed at a particular state, and vice versa (Woodward, 2015b; 2016). For example, suppose *Alcohol* = 1 or 0 represents whether one has a habit of alcohol use or not; *Cancer* = 1 or 0 represents whether one develops lung cancer or not. *Alcohol* and *Cancer* are conceptually and ontologically distinct in the sense that every combination of their values is attainable under some hypothetical intervention. Therefore, the dependence relation between them is causal (rather than non-causal). Compare this with the relation between *Speak* = 1 or 0 (whether one speaks or not) and *SpeakLoudly* = 1 or 0 (whether one speaks loudly or not). It is impossible to attain *Speak* = 0 while *SpeakLoudly* = 1. This is because the value of *SpeakLoudly* non-causally depends on the value of *Speak* such that *Speak* = 0 conceptually precludes the possibility of *SpeakLoudly* = 1.

Clearly, a causal structure and the correlations it generates are not fully distinct or independently manipulable in this sense. For instance, assuming that the causal structure *Mutation* → *Cancer* ← *Alcohol* satisfies CMC and CFC, it is neither conceptually nor ontologically possible for *Mutation* and *Alcohol* to be marginally dependent in the overall population while simultaneously conditionally independent among individuals with lung cancer. This demonstrates that the dependence relation

between the relevant features of this causal structure and the observed correlations it generates is modally stronger than what a causal relation allows.[22]

Given that the explanandum non-causally depends on the explanans in a causal-structure explanation, if the relation view is correct, causal-structural explanations should count as a kind of non-causal explanation. The explanans view and the relation view thus disagree about whether causal-structural explanations are causal explanations. Philosophers need to take a stand on which of the two views more accurately draws the boundary between causal and non-causal explanations.

6. Conclusion

Causal-structural explanations are widespread in scientific practice yet remain underexplored in philosophy, even though philosophers familiar with causal modelling may have recognized them. In this chapter, we have offered a systematic analysis of such explanations and argued that the way in which the explanans gives rise to the explanandum is not itself causal in nature.

We take this to show that the received conception of causal explanation is inadequate, because it is ambiguous as to whether it is the explanans, or the dependence relation between the explanatory relata,

[22] Relatedly, mechanistic explanations of probabilistic correlations or causal regularities display an explanatory pattern similar to that of causal-structural explanations. On the one hand, the underlying mechanism provides causal information about why the explanandum correlation or regularity obtains. On the other hand, the relation between a mechanism and the correlation or regularity it generates is often regarded as constitutive rather than causal, although this is not uncontroversial (see Craver, 2007; Baumgartner and Gebharter, 2016; Craver et al., 2021).

that distinguishes causal explanations from non-causal ones. A satisfactory conception of causal explanation must specify which of the two is the defining feature of causal explanation. If it is the explanans, then causal-structural explanations are causal explanations. If, instead, it is the dependence relation, then they are non-causal explanations.[23]

Acknowledgements

Chapter 3, in full, is a reprint of the paper co-authored with Zili Dong, which is accepted for presentation at PSA 2026 and conditionally accepted for publication in *Philosophy of Science*. The dissertation author was the corresponding and co-first author of this paper.

[23] In our view, the explanans view is more plausible. We will not offer a full defense here but simply note that scientists typically regard the search for appropriate causal structures to account for observed correlations as a key practice of causal inference. The explanans view is preferable because it better accommodates this feature of scientific practice.

Chapter 4

Simpson's Paradox Beyond Confounding

Co-authored with Zili Dong and Shimin Zhao

1. Introduction

Simpson's paradox (SP), in its most striking form, is a phenomenon where a statistical association between two variables X and Y in the entire group (or aggregate data) reverses in every sub-group (or disaggregate data). SP also includes cases where the association between X and Y in the entire group disappears in each sub-group, as well as cases where an association between X and Y appears in each sub-group even if they are unassociated in the whole group. Since we typically partition a group based on some third variable of interest Z (e.g., *Sex*), we may also say that SP occurs when the association between X and Y *reverses*, *disappears*, or *emerges*, after conditioning on Z. For a recent survey of SP, see Sprenger and Weinberger (2021).

Two clarifications on the above definition of SP are needed. Firstly, following Hoover (2003) and Sprenger and Weinberger (2021), we understand SP primarily as a sample-level phenomenon that can be readily observed in statistical data. In this chapter, the concept of sample *association* is distinguished from the concept of population (or probabilistic) *correlation*. Of course, if we have a case of SP in which

sample associations between X, Y, and Z adequately indicate the probabilistic correlations between them, this will be a case of SP defined in terms of both association and correlation.[24] Secondly, even though we agree that causality plays a key role in understanding many important cases of SP, the definition of SP we adopt in this chapter does *not* stipulate that SP must be a causal phenomenon.[25]

Probably the most well-known example of SP is the case of graduate admissions at the University of California, Berkeley in 1973 (Bickel, Hammel & O'Connell, 1975). It was recorded that at the university level, about 44 percent of the males and about 35 percent of the females were admitted. This means that being female was negatively associated with being admitted to the University of California, Berkeley, which suggests that there might have been discrimination against female applicants in the admissions process. However, if we break down the data, we will find that, in the majority of the departments, there was no significant bias against female applicants. In fact, in a few departments, females were even more likely to be admitted than males. This poses the question of whether there was truly sex discrimination that affected the admissions committee's decisions. That is, if we want to

[24] In most cases, we can safely ignore the difference between sample associations and probabilistic correlations. Still, there are cases in which it is important that we separate them, as we shall see later in the chapter.

[25] It has been suggested that SP, in its nature, is a *causal* phenomenon (e.g., Pearl, 2014; we shall come back to this later in the chapter). For Pearl, a genuine case of SP must be embedded in a causal context. Apparent cases of SP that lack causal context are dismissed by him as not genuinely paradoxical (he calls such cases "Simpson's reversal"). Although we think the distinction Pearl draws between Simpson's paradox and Simpson's reversal is well-motivated, we also find it somewhat *ad hoc* to stipulate that SP must be a causal phenomenon. We show in Section 3 that there are more principled reasons why cases of SP that lack causal context should be distinguished from those cases that have a causal context, without having to draw the distinction by stipulation.

identify the existence of sex bias in the admissions process, should we look at the university-level data or the department-level data?

Although SP may seem like a purely statistical or probabilistic oddity, philosophers have long recognized its causal roots. Back in the 1980s, SP was posed as an important challenge to probability-raising accounts of causality (see, e.g., Cartwright, 1979). In this context, Cartwright rightly pointed out that the association between the cause-variable X and the third variable Z is essential for the occurrence of SP.[26] To avoid SP, she proposed that we measure causal effects relative to the so-called "causally homogeneous" populations or reference classes in which there is no association between X and Z, since Z can be seen to have been "held fixed" in such homogeneous reference classes. However, this proposal, embedded in the probability-raising approach to causality, is subject to the approach's inability to explicitly represent causal structures underlying the SP. As early as 1987, Irzik and Meyer had realized the inadequacy of Cartwright's solution and suggested we analyze the causal structure of SP using tools of causal modelling (Irzik & Meyer, 1987).[27] The tool they used is the method of path analysis (invented by Sewall Wright around 1920), which is a precursor to the more powerful framework of graphical causal modelling developed later by Spirtes et al. (2000) and Pearl (2009).

[26] See also Sprenger and Weinberger (2021) for a detailed explanation of why an association between X and the third variable Z is *necessary* for SP to occur.

[27] Irzik and Meyer's (1987) analysis of SP is remarkably farsighted; its core idea is basically the same as Pearl's confounding-based analysis of SP (see the illustrative example they give on p. 513).

Proponents of the framework of graphical causal models propose to analyze SP in *causal-graphical* terms (Spirtes et al., 2000; Pearl, 2009; 2014; Pearl et al., 2016; Pearl & Mackenzie, 2018). Notably, Pearl et al. (2016) assert that they can "fully resolve Simpson's Paradox by determining which variables to measure and how to estimate causal effects under confounding" (p. 44). For Pearl et al., *confounding* is to be analyzed in terms of causally interpreted directed acyclic graphs (DAGs). Essentially, their proposal is that for *all* types of SP involving X, Y, and Z, the paradox can be resolved by construing the partitioning variable Z as a confounding variable relative to a causal DAG over $\{X, Y, Z\}$. We will explain this in more details in Section 2 (see especially Figure 4.1 there).

In this chapter, we contend that this causal-graphical analysis of SP, despite offering genuine insight, is not complete and cannot fully resolve SP as Pearl et al. have claimed. We acknowledge that many important types of SP do arise from confounding; however, we do not think this is a universal feature of SP. For one thing, some cases of SP lack causal context, as has been argued by Bandyopadhyay et al. (2015). While acknowledging that Bandyopadhyay et al. raise a logically valid objection against the completeness of Pearl's causal-graphical analysis of SP, we will examine their argument from the perspective of scientific and statistical practice and show that, at least from this perspective, it poses no significant challenge to Pearl's analysis. More importantly, we show that even if we narrow down our attention to cases of SP that do need a causal treatment, we can still find various types of SP that do *not* involve confounding of any sort. After carefully examining these types of SP, we conclude that SP should be seen as a symptom with many aetiologies and there appears to be no unified analysis that can capture all of them.

The rest of the chapter proceeds as follows. In Section 2, we first present Pearl's causal-graphical analysis of SP and then point out its limitations which are to be further examined in subsequent sections. In Section 3, we argue that Bandyopadhyay et al.'s alleged counterexample to Pearl's analysis (i.e., the Marble example) does not constitute a threatening objection to Pearl, because the example presupposes strong accidental associations which are unlikely to be encountered in ordinary life or scientific practices. Section 4 discusses three cases of SP which all require a causal analysis, but Pearl's analysis does not apply. Specifically, in Section 0, we discuss a case of SP arising in an evolutionary context due to accidental associations (random genetic drift). Section 0 discusses a case of SP arising from the use of inappropriate aggregate variables as causes, and Section 0 discusses a case of SP arising from inter-unit causation (illustrated using a non-stationary time series example). Section 5 extends the discussion to a generalized version of "SP-type" phenomena known as the amalgamation paradox (AP).[28] It has been recognized by epidemiologists—but less known by philosophers—that some cases of AP can occur without confounding; for this reason, we note that Pearl's analysis of SP cannot be generalized to AP. Section 6 is a brief conclusion.

[28] It is sometimes said that AP is "the most generalized version of [SP]" (Sprenger & Weinberger, 2021), and sometimes SP and AP are used as synonyms (as in, e.g., Hernán, Clayton & Keiding, 2011). Admittedly, this use of terminology is confusing. To avoid confusion, we make a clear distinction between SP and AP in this chapter: AP is defined as a broader category than SP, with SP being a special case of AP. This distinction will be important for our discussion in Section 5.

2. The Scope and Limitations of the Causal-Graphical Analysis

In this chapter, we focus on Pearl's (especially, 2014) causal-graphical analysis of SP since, to our knowledge, this is the most influential and systematic treatment of SP by far. At the centre of his analysis is the framework of graphical causal models. The first thing to note is that although Pearl's analysis of SP is often referred to as the "causal analysis," it is more accurate to call it the "causal-*graphical* analysis." This is not merely a verbal issue: while graphical modelling is undoubtedly an important tool in causal inference, a graphical model by no means captures everything interesting about a system's causal properties (Cartwright, 2001; Dawid, 2010).

Directed acyclic graphs (DAGs) are the most often used type of graphical models in causal inference. A causally interpreted DAG G consists of a set of vertices, which represent a set of causal variables $\mathbf{V} = \{X, Y, Z, ...\}$, and a set of edges, which represent *direct* causal relations between the variables. X is a direct cause of Y (relative to G) in the sense that it is possible to change the value of X through some atomic or ideal interventions such that the probability distribution of Y will change accordingly, when all other variables in the graph are held fixed by interventions (Woodward, 2003; Pearl, 2009). G is directed, which means that all the edges are single-headed arrows. We define a causal path between X and Y on G as a path on which the edges between X and Y are all directed in the same direction (e.g., $X \rightarrow Z \rightarrow W \rightarrow Y$). G is also acyclic, which means that it contains no causal circle (i.e., causal path that starts and ends with the same variable, e.g., $X \rightarrow Z \rightarrow X$). Besides, G and its corresponding joint probabilistic distribution over $\mathbf{V}$ are assumed to satisfy the *causal Markov condition* (a modern successor

of Reichenbach's principle of common cause). This condition says that for any variable X in $\mathbf{V}$, X is probabilistically independent of every other variable except X's descendants (i.e., X's effects), conditional on X's parents (i.e., X's direct causes).

DAGs have proven empirically fruitful in representing and analyzing the causal structure or data-generating process relevant to a causal investigation. They are particularly suited for handling the notorious problem of *confounding* in causal inference (Shrier & Platt, 2008; Pearl et al., 2016).[29] The primary task of causal inference in many scientific domains (especially in high-level sciences such as biology, sociology, epidemiology, etc.) is to identify and estimate causal effects. Confounding is an important type of systematic source of error that might occur during causal effect estimation. An error is systematic means that its occurrence is not accidental; that is, it does not occur by chance. For example, if we use the marginal association between X and Y in the observational data to measure the direct effect of X on Y, we may misestimate the effect when there are indirect causal paths between X and Y that bring about indirect effects of X on Y. For Pearl, such kind of discrepancy should be assessed using DAGs which can visually represent all the relevant causal assumptions. With these causal assumptions and the help of certain graphical rules (e.g., the back-door criterion), we can then eliminate confounding by adjusting for or conditioning on a (sufficient) set of confounding variables (Pearl et al., 2016).

[29] For various reasons, the term "confounding" has been used in confounding ways (pun intended). For example, sometimes "confounding" is used to refer merely to bias due to lurking common causes (cf. Hernan et al. 2011). Our usage here is much broader, following Pearl, Glymour, and Jewell (2016). On this broad usage, a confounding variable may also be a collider or a mediator; see our discussion below.

For our purposes below, it suffices to focus on the estimation of the direct effect of X on Y relative to a pre-specified DAG on $\{X, Y, Z\}$. Relative to a DAG on $\{X, Y, Z\}$, our estimation of the direct effect of X on Y will be confounded if and only if any of the following situations obtains: (a) we fail to adjust for Z when Z is a common cause of X and Y, (b) we mistakenly condition on Z when it is a common effect of X and Y (here Z is also called a "collider" and this type of confounding is also known as "collider bias"), or (c) we fail to adjust for Z when Z is a mediator between X and Y (i.e., Z is on an indirect causal path between X and Y). Figure 4.1 illustrates these three cases of confounding with DAGs.

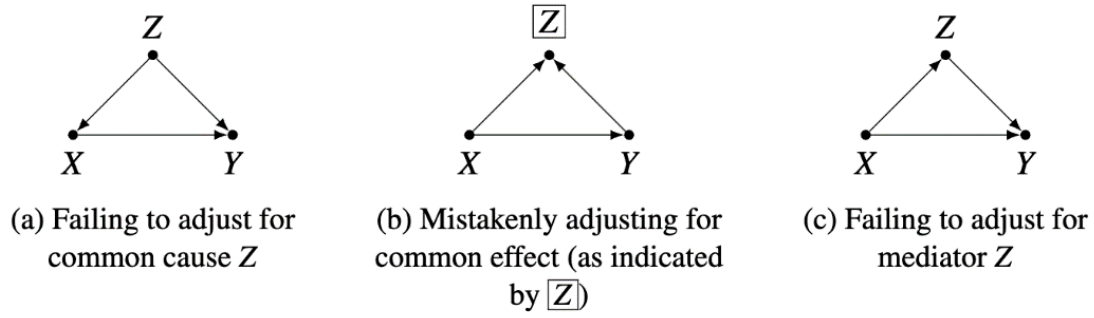


Figure 4.1 Three possible cases of confounding in the estimation of the direct effect of X on Y.

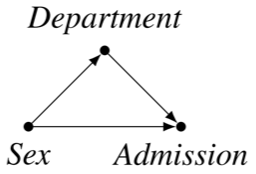
Consider a concrete example. If we are interested in the effect of Paxlovid pills in reducing deaths from COVID-19, we cannot simply compare the death rate in a group of COVID-19 patients who took the pills with the death rate in another group who did not. This estimation of the drug's effectiveness is confounded since patients at a higher risk of dying from COVID-19 are also much more likely to receive

the pills. In other words, risk factors for COVID-19 such as age or chronic diseases, if unadjusted for, will lead to confounding (of type (a) mentioned above) in the effect estimation.

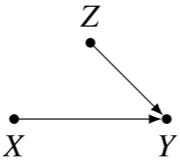
Pearl (2014) suggests that we use DAGs to analyze causal structures that are at work behind various cases of SP. In his analysis, SP is diagnosed as a peculiar consequence of confounding. The basic idea is the following. The reason we find SP "paradoxical" is that we think how X (e.g., *Sex*) affects Y (e.g., *Admission*) should not depend on the level at which the influence is measured. If the associations between X and Y in the large group and in the sub-groups disagree with each other, they cannot both indicate the "true" effect of X on Y—at least one of the associations must be "spurious." For it violates our causal intuition to say that depending on the way we look at the data, the applicant's sex can both influence, and not influence, admissions. Pearl's key insight is that the kind of spurious association that leads to SP should be seen as resulting from confounding. If we can eliminate the confounding responsible for a case of SP (assuming that the relevant causal structure is known), we will obtain an unconfounded estimation of the true effect of interest. The paradoxicality and counter-intuitiveness of SP will then be explained away.

Pearl (2014) identifies a group of causal structures that can give rise to SP, together with a group of causal structures that cannot. Here we consider one example representative of each type. First, consider the DAG in Figure 4.2a, which depicts the putative causal structure responsible for the example

of Berkeley's graduate admissions.[30] Note that there are two causal paths from *Sex* to *Admission*, which means that, relative to this DAG, *Sex* is both a direct and an indirect cause of *Admission*. This type of causal structure is prone to generating SP, because the causal influence that *Sex* has on *Admission* along the direct causal path and the indirect one could be in opposite directions.



(a) The causal structure for Berkeley's graduate admission example



(b) A causal structure that (allegedly) cannot generate SP, according to Pearl

Figure 4.2 A comparison of a causal structure that can generate SP and a causal structure that (allegedly) cannot generate SP.

More specifically, *Sex* will have a direct effect on *Admission* if sex bias does exist in the admissions process (e.g., being female causes one to be discriminated against in this process).[31] At the same time, one's sex may influence one's department choice, which can further affect one's chance of being admitted

[30] This causal graph is adapted from Pearl and Mackenzie (2018, p. 312, Fig. 9.4). Pearl and Mackenzie also considered more complicated causal structures, but these complications are not essential for our discussion here.

[31] Note that the direct effect of *Sex* on *Admission* only reflects possible sex biases during the admissions process. This cannot tell us anything about structural sexism.

by the university because some departments are harder to get in than others. Thus, *Sex* also has an indirect effect on *Admission*, which is mediated by *Department*. In this case, the association between *Sex* and *Admission* in the university-level data, without adjusting for *Department*, gives us an estimation of the sum of the direct and indirect effect (i.e., the *total* effect) of *Sex* on *Admission*. To examine whether there is sex bias against female applicants—that is, whether *Sex* has a *direct* effect on *Admission*—we should instead look at the association between *Sex* and *Admission* when *Department* is adjusted for, given that the decision to admit an applicant was processed within each department. In other words, we should rely on the association between *Sex* and *Admission* in the department-level data to infer whether female applicants were discriminated against in the admissions process. If one tries to identify sex discrimination using university-level data, the estimation will be confounded.

In contrast, the causal structure in Figure 4.2b is claimed by Pearl to be unable to give rise to SP. His reason is that, in this DAG, Z cannot bring about any of the three aforementioned types of confounding: Z is not a common cause, a collider, or a mediator. Given the causal Markov condition, the DAG implies that X and Z are *probabilistically* independent. Under the assumption that probabilistic independence implies statistical independence, X and Z should be found unassociated in the data. It is this assumed absence of association between X and Z that justifies Pearl's claim that SP will not arise for this DAG. However, this assumption does not always hold. It is still possible that X and Z, despite being causally and probabilistically independent, are *accidentally* associated in the data we collected, especially when the sample size is small; in such cases, we may still encounter SP in the absence of confounding. Fortunately, this type of situation is unlikely to arise in sufficiently large samples.

Additionally, note that both associations arising from chance and associations resulting from confounding are "spurious" in the sense that they do not reflect true causal effects. The difference between them, however, is that in statistical practices, the latter is much more robust and common. This is also why Pearl simply disregards the possibility of accidental associations in his discussion. We will come back to these points in Section 3.

Undoubtedly, Pearl's causal-graphical analysis of SP contains genuine insight and covers a broad range of cases of SP. That said, we believe this analysis is inadequate in important ways. His analysis focuses on those cases of SP in which the associations involving X, Y, and Z are supposed to be explained by positing an underlying causal structure over $\{X, Y, Z\}$, represented using a DAG. However, it is well-known that DAGs make certain assumptions about the causal structures they represent which do not necessarily hold in practice.[32] In particular, a standard DAG presupposes at least two things. Firstly, the causal variables have been well-chosen. Poor choices of variables, for instance, using variables that are inappropriately defined or logically connected, may bring trouble to causal inference (see Woodward, 2016). Secondly, an association between two variables can always be causally explained, either as a result

[32] Broadly speaking, our point here aligns with Spanos' claim that Pearl's "causal explanation of the paradox largely ignores some of these empirical issues by viewing it as a purely probabilistic conundrum ... when models are estimated using actual data, one needs to secure the validity of the model assumptions vis-a-vis the data before any causal information can be utilized reliably" (Spanos, 2021, p. 608). In other words, investigators run the risk of introducing unwarranted causal information into the model when those model assumptions are not examined for their validity.

of genuine causation or as that of confounding.[33] As we shall see later, these assumptions, while making DAGs expedient to use, may not always hold.

It follows that Pearl's causal-graphical analysis of SP will not apply in either of the following two types of cases:

> **(Case 1)** The occurrence of SP does not require the existence of any causal relationships among the relevant variables (i.e., *X*, *Y*, and *Z*). That is, none of the associations between *X*, *Y*, and *Z* in this case of SP needs to be generated by causation among these variables. To make sense of this type of SP, causal information is irrelevant.

> **(Case 2)** SP occurs in a context in which the effect of *X* on *Y* is queried, prompting an investigation into the causal story underlying SP. Moreover, this causal story is required for explaining away SP in this case. However, the presence of this type of SP does not result from confounding; its root lies somewhere else.

The sort of counterexample to Pearl's analysis in **Case 1** has been extensively examined by Bandyopadhyay et al. (2015), which will be the focus of our discussion in the next section. As Bandyopadhyay et al. argue, at least in some cases, SP need not be analyzed in causal terms. We agree with them that their counterexample shows that Pearl's analysis cannot adequately account for *all* possible

[33] It must also be acknowledged that these assumptions may be made by other approaches to causal inference as well. For example, that causal variables should be in some sense well-chosen is a prerequisite not just for graphical causal modelling but also for the potential-outcomes approach to causal inference.

instances of SP. Nevertheless, we believe it is equally important to explore counterexamples in the line of **Case 2**. For one thing, while Bandyopadhyay et al.'s counterexample, as an instance of **Case 1**, constitutes a *logically* valid objection to Pearl's analysis, we believe that it is not a *practically* strong challenge. As we shall explain, the occurrence of their counterexample is extremely rare in ordinary and scientific contexts. In our view, it is counterexamples in the line of **Case 2** that offer a stronger ground against the adequacy of Pearl's analysis.

For another, **Case 2** has received limited attention in the literature, although the question of whether there are causal grounds of SP that do not involve confounding and thus are not subject to Pearl's analysis is evidently important. Pearl's causal-graphical analysis regards confounding as an indispensable condition for SP and considers confounding adjustment as sufficient for resolving SP. This, if true, implies that practitioners of causal inference need not worry about possible instances of SP if no confounding is present. In this chapter, we show that this is not the case. SP can arise in causal contexts where confounding is absent. In Section 4, we offer three cases of SP that fall under **Case 2**. As we shall see, for all three cases, although causal information plays a key role in explaining away the paradox, none of them involves confounding.

3. Simpson's Paradox Without Causation? A Statistical-Practice Perspective

Bandyopadhyay et al. (2011; 2015) (and more recently, Sarkar & Bandyopadhyay, 2021) defend a non-causal, "*logic-based*" analysis of SP. According to them, at least in some cases, SP is fundamentally an arithmetic oddity, whose nature has nothing to do with causality. In their view, SP "involves the reversal of the direction of a comparison or the cessation of an association when data from several sets are pooled" (Bandyopadhyay et al., 2015, p. 13; note that this is essentially equivalent to the definition of SP we give at the beginning of this chapter). Importantly, Bandyopadhyay et al. emphasize that the satisfaction of these conditions need not presuppose any causal relations among X, Y, and Z. Moreover, causal information is also unnecessary for explaining why data patterns satisfying these conditions can give us a feeling of puzzlement (and thus be regarded as "paradoxical").

Therefore, Bandyopadhyay et al. claim, it is not the case that SP can only be explained away by positing causal relations among the relevant variables. Note that they do not deny that sometimes we need to analyze SP causally. However, "SP has to do with causality *only if* we ask the *what-to-do* question"—the question of "What should one *do* when confronted with a typical case of the paradox?" (Bandyopadhyay et al., 2015, p. 14; emphasis added). In other words, the causal structure underlying an instance of SP becomes relevant only in circumstances where one is considering what decision should be made in view of the associations exhibited in this instance of SP. For example, when we encounter a situation where it appears that the effect of a treatment reverses after conditioning on sex, the question

of how such "contradictory" evidence should guide the use of the treatment arises. In cases where no such decision-making concern arises, according to Bandyopadhyay et al., there would be no need to understand SP through a causal lens.

Table 4.1 The Marble example with fictitious data (adapted from Bandyopadhyay et al., 2015, Table 5).

Red Marble Rates	Bag 1	Bag 2	Total
Big Marbles	180/200 = **90%**	100/300 = **33%**	280/500 = 56%
Small Marbles	480/600 = 80%	10/100 = 10%	490/700 = **70%**

Bandyopadhyay et al. use the Marble example to illustrate their point. Suppose there are two bags of marbles with different sizes and colors. Table 4.1 reports how many marbles of a specific size-color pair there are in each bag. It is observed that within both bags (Bag 1 and Bag 2), big marbles are more likely to be red, compared with small marbles. Given this, it seems reasonable to expect that the same pattern will hold once we merge all the marbles together. However, this expectation cannot be fulfilled: the association between *Size* and *Color* reverses in the aggregate data. We find this result surprising, peculiar, or perplexing because it violates our expectation. According to Bandyopadhyay et al., "[t]here are no causal assumptions made in this example" regarding how *Size*, *Color*, and *Bag* are related to one another (Bandyopadhyay et al., 2015, p. 19). This implies that no confounding can be appealed to in accounting

for why this instance of SP occurs. Granted that this is the case, Bandyopadhyay et al. contend that at least in this example, SP has a purely arithmetic root, which implies that "SP is not basically causal" (p. 13). Thus, Pearl's causal-graphical analysis cannot be a universal treatment of SP.

We agree that Bandyopadhyay et al.'s Marble example is a genuine case of SP: the example does involve the kind of association reversal characteristic of SP, and therefore, conforms to the definition of SP we give at the beginning of this chapter.[34] In addition, we agree that the example does stimulate our intuitive perplexity and that the perplexity can be explained away without invoking any causal postulates involving the relevant variables. Indeed, the example is designed in this way in order to constitute a counterexample to Pearl's causal-graphical analysis.

Nevertheless, we find the Marble example inadequate in an important sense. From a *statistical-practice* perspective (as opposed to a purely "logic-based" or "arithmetic" one), this example presupposes a highly uneven distribution of marbles in the two bags. In particular, in Bag 1, ¾ of the marbles are small, whereas in Bag 2, ¾ of the marbles are big, which means there is a *strong association* between *Size* and

[34] Pearl would give a different response to the Marble example: he would think that this is not a genuine case of SP at all but merely a case of Simpson's "reversal" (see footnote 25). For Pearl, it is essential for a genuine case of SP that it violates our *causal* intuition. If we understand him correctly, he seems to think that the violation of the causal intuition (as we talked about in Section 2) should be a *defining* feature of SP. Although we agree with Pearl that cases of SP that violate this causal intuition are the exemplars of SP, and we also agree on the explanatory significance of this causal intuition, we hesitate to make this feature definitional of SP. After all, the definition of SP we give in Section 1—which is also widely used in the literature—is not formulated in causal terms. Besides, many of our concepts have both typical and atypical cases. The Marble example seems to be better conceived as a less typical case of SP.

Bag. This distribution of the marbles is extremely rare from a statistical practice perspective. This is because, assuming that the causal Markov condition is satisfied, these variables are not expected to be found highly associated in almost all the data we may encounter in practice (given the relatively large sample size in this example) unless we posit causal connections among them. In fact, under the hypothesis that there is no causation (and thus no correlation) between *Size* and *Bag* in the Marble example, the probability of observing a difference in the proportions of big marbles between the two bags is 50% (i.e., 300/400 – 200/800) or more is lower than 0.000001: it happens less than once in a million times. Similarly, the strong association between *Color* and *Bag* presented in Table 4.1 is also extremely rare, assuming that *Color* and *Bag* are not correlated.

Note that Bandyopadhyay et al. cannot explain the strong association between *Size* and *Bag* and that between *Color* and *Bag* in Table 4.1 by appealing to a biased process of disproportionally sorting more small marbles and red marbles into Bag 1. This is because, in that case, the Marble example would require a *causal* analysis, since this biased sampling process is actually an instance of collider bias (see Section 2 for more on collider bias). That is, if the size and color of a marble affect which bag it will be put into, then *Bag* will become a collider whose value is affected by both *Size* and *Color*. Now that a causal structure behind the data has been introduced, the Marble example will no longer be a counterexample to Pearl's causal-graphical analysis.

Therefore, the associations in the Marble example can only be "chancy" or *accidental* in the sense that they are a result of very rare random fluctuations in the process of sorting marbles into bags. These accidental associations, especially the one between *Size* and *Bag*, play a key role in generating SP in this

example: if we were to render the proportions of big marbles in both bags roughly the same, SP would not occur. The fact that the occurrence of SP in the Marble example depends on strong accidental associations implies that, *statistically* speaking, this kind of SP (**Case 1**) is far less common or robust, compared to typical cases of SP such as the Berkeley admissions example. We have very little reason to expect that data patterns similar to the Marble example will be ever observed in statistical practice. By contrast, in the Berkeley example, there is a causal structure that *robustly* generates associations among *Sex*, *Department*, and *Admission*. This causal structure may also be found in other years or places. For this reason, we may expect to find similar occurrences of SP in the admissions processes in the years after 1973 at UC Berkeley or other universities.

We think Bandyopadhyay et al.'s discussion of SP does not give due emphasis to the fact that instances of SP generated by robust causal structures and those generated by chancy fluctuations are *not* on a par with respect to their importance in scientific practice. In most contexts, scientists collect and analyze data in order to reveal underlying causal structures (or for other causally relevant goals, including explanation, prediction, understanding, and control).[35] The connection between a causal structure and

[35] Notably, there are data-generating processes that are non-causal but remain scientifically significant. For example, nonlocal quantum correlations that violate Bell inequalities are usually considered not subject to a causal interpretation (cf. Myrvold et al., 2024). Despite this, these correlations have been robustly observed in well-designed Bell experiments, which is why they are of great scientific significance. See Frisch (2020) for a brief survey of this issue; see Wood & Spekkens (2015) and Näger (2022) for recent discussions. However, because the literature on SP has almost exclusively focused on high-level sciences where quantum effects can be safely ignored, we will also set aside quantum correlations in this chapter. We thank an anonymous reviewer for suggesting this point.

the associations it robustly generates allows scientists to infer the former from the latter. However, when associations seem to provide "contradictory" evidence for the underlying causal structure, as we saw in the Berkeley admissions example, it can be particularly puzzling; this is why SP as a statistical phenomenon has received so much scientific attention.

Importantly, since accidental associations may mislead causal inference, scientists will take various measures (e.g., collecting more data, and conducting significance tests) in attempts to reduce the possibility that the evidence they observe comes from an accidental association in a particular dataset. These measures enable scientists to quickly discover and discard, to their best knowledge, accidental associations due to chancy fluctuations in sampling. For this reason, SP arising from accidental associations is of minimal scientific relevance.[36] Therefore, we claim, it is mainly through a causal lens that scientists come to see cases of SP as intellectually intriguing and practically significant. In contrast, although the Marble example successfully demonstrates that Pearl's causal-graphical analysis is not a universally valid treatment of SP, its dependence on an extremely rare random fluctuation precludes it from being a scientifically and practically significant counterexample.

[36] Note that we are not denying the value of investigating "outliers" or "exceptional" data points. As pointed out by an anonymous reviewer, there are ample examples in medicine where paying attention to "exceptional responses" to a treatment in clinical trials led to important medical progress (Mukherjee, 2015). However, as Mukherjee also made clear, the rarity of these cases is due to complex interactions between numerous *causal* factors. These factors are of scientific interest because if we can discover and control these factors, we will be able to robustly generate these "exceptional" cases. Therefore, they are not "chancy" fluctuations in the statistical sense which are not generated by any interesting causal factors.

Lastly, we disagree with Bandyopadhyay et al. (2015) that causality only matters when decision-making questions are asked regarding a case of SP. One may still want to conduct a causal inquiry about how an instance of SP is generated when one's goal is purely *epistemic,* such as explanation or understanding. It is quite often that people scrutinize data merely for the sake of gaining an understanding of its underlying data-generating process, without immediate concern for decision-making. For instance, one might have a purely epistemic interest in the historical question of whether there was truly sex bias in Berkeley's 1973 graduate admissions process, even if she does not want to make any decisions.

Interestingly, we do find a type of SP that arises due to an accidental association between X and Z, and at the same time, has a more visible scientific significance. Unlike the Marble example, however, this new type of SP needs to postulate causal relations among the relevant variables. So, these two types of SP, despite both relying on accidental associations, are importantly different: one is situated in a causal context whereas the other need not be. This is why we will leave the discussion to Section 4.

4. Simpson's Paradox with Causal Contexts but Beyond Confounding

This section aims to demonstrate that even among cases of SP that are embedded in a causal context, some of them cannot be analyzed as a consequence of confounding. Note that we are not saying that these cases should not be analyzed causally, nor are we saying DAGs are no longer useful for analyzing them. Our claim is specifically that these cases need to be treated in a careful and nuanced manner which goes beyond what Pearl's (2014) analysis can offer.

4.1 Accidental Associations

The previous section discussed a form of SP free of causal context, as illustrated using the Marble example. In this section, we shall see that similar to the Marble example, accidental associations can also play a key role in generating SP when the instance of SP is situated in a causal context. More specifically, an accidental association between X and Z can generate SP over $\{X, Y, Z\}$ when they form a causal structure like the following: $X \rightarrow Y \leftarrow Z$ (see Figure 4.2b). Recall that in Section 2, we noted that this causal structure cannot generate SP, *only* on the assumption that there is no accidental association between X and Z. But if X and Z are, in fact, accidentally associated, SP can still arise in this causal structure.

Consider the following example (see Table 4.2; adapted from Sober, 2024, p. 44) about how being selfish or altruistic may affect the expected number of offspring of an individual (i.e., fitness). There are two groups, A and B, each containing two traits: being selfish, or being altruistic.[37] A trait's fitness in a group is defined by how many offspring individuals of the trait in that group produce on average. For example, in group A, there are 200 selfish individuals, and they have 800 offspring in total, so the fitness of selfishness in group A is 800/200=4. As shown in the table, being selfish is positively associated with fitness in both group A and group B. However, the association between selfishness and fitness reverses

[37] For the sake of simplicity, let us assume asexual reproduction, no mutation, and no migration.

once the two groups are analyzed as a whole (i.e., as a metapopulation). This is paradoxical if we interpret fitness as representing the *causal* propensity of a trait in producing offspring.

Table 4.2 A hypothetical example of SP involving the fitness of being selfish or altruistic.

Fitnesses	Group A	Group B	Total
Selfish	800/200 = **4**	1600/800 = **2**	2400/1000 = 2.4
Altruistic	2400/800 = 3	200/200 = 1	2600/1000 = **2.6**

The paradoxicality of this example is primarily due to the existence of a very strong association between variables *Selfish* and *Group*: in our example, group A is dominantly altruistic whereas group B is dominantly selfish. Without this association, SP would not be possible. But where does this association come from? It might be because having a particular trait *causes* one to be in a certain group—making this case of SP similar to the Berkeley example.[38] But what we want to show below is that the association need not necessarily come from a causal source.

[38] This situation also seems to be what Sober (2024, Sect. 3.2) has in mind.

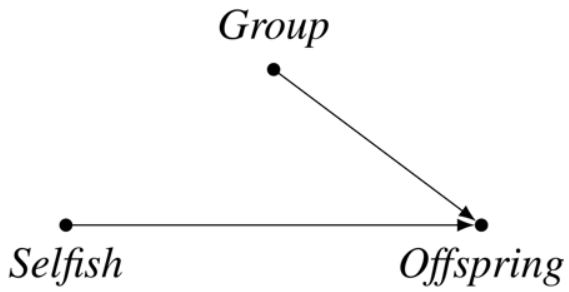


Figure 4.3 A causal graph for the SP involving *Selfish*, *Group* and *Offspring*, without positing a causal relation between *Selfish* and *Group*.

Even if we assume that *Selfish* and *Group* are causally *in*dependent (as depicted in Figure 4.3**Figure 4.*3*)**, we may still be able to accidentally observe a strong association between them. Regarding our example specifically, such a strong association may have resulted from genetic drift, such as the founder effect (Dobzhansky & Pavlovsky, 1957). That is, by pure chance, two relatively small groups of individuals may happen to have highly uneven distributions of altruism and selfishness upon being separated from the larger group. Put differently, the association between *Selfish* and *Group* in the above example can be seen as a result of random fluctuations.

Clearly, this chancy association is not a result of confounding, since *Selfish* and *Group* have been assumed to be causally independent. Still, this accidental association can mislead causal effect estimation and lead to SP. Therefore, if we are to estimate the fitness of selfishness reliably, we need to first make

sure that genetic drift does not generate a significant accidental association between *Selfish* and *Group* in the data.

4.2 Aggregate Variables

An important type of scenario that can supply "paradoxical" associations needed for SP but cannot be analyzed away in terms of confounding involves the use of aggregate or summed variables as causes. An aggregate variable is one that can be written as the sum of two or more other variables (e.g., $X=X_1+X_2$). It is known that an aggregate variable may have an *ambiguous* effect on an outcome of interest, if the variables it sums up have heterogenous effects on the outcome (Spirtes & Scheines, 2004). A well-known example of such kind of aggregate variable is total cholesterol (*TC*). *TC* consists of both low-density lipoprotein (*LDL*) and high-density lipoprotein (*HDL*), which have opposite effects on cardiovascular diseases (*CVD*). The ambiguity in the effects of such kind of aggregate variables, as we shall see, opens a door for SP.

Imagine that we have a group of patients in which levels of total cholesterol are *unassociated* with the incidence of cardiovascular diseases, which suggests the prima facie result that *TC* has no effect on *CVD*. However, if we condition on levels of *HDL*, surprisingly, *TC* becomes *positively* associated with *CVD* (i.e., a higher level of *TC* is associated with a higher *CVD* rate, conditional on *HDL*). So, a positive association emerges upon conditioning the third variable—making this a case of SP. Table 4.3 represents the results of a fictitious dataset for this example. It shows that *CVD* is unassociated with levels of *TC* in the entire study group (50% vs. 50%). However, within both sub-groups, having a high level of *TC* seems

to make the patients more likely to develop cardiovascular disease. How should we explain this paradoxical result?

Table 4.3 Fictitious data for the incidence of cardiovascular disease among sub-groups (with low and high levels of *HDL*) and among the whole group.

CVD Rates	Low HDL	High HDL	Total
Low *TC*	40/70 = 57.1%	20/50 = 40%	60/120 = 50%
High *TC*	80/130 = **61.5%**	120/270 = **44.4%**	200/400 = 50%

Due to the *non-causal* (specifically, logical) relationship among *TC*, *HDL*, and *LDL* (i.e., *TC=LDL+HDL*), including these three variables in a single *causal* graph may cause trouble. So, to say the least, caution is needed if one attempts to provide a graphical analysis of the above example. In particular, it is helpful to represent non-causal relationships using dashed arrows so as to distinguish them from causal relationships (see Figure 4.4). Besides, in the DAG we draw, there is no need to draw a causal arrow from *TC* to *CVD*, given that all the work *TC* appears to be doing is in fact done by *LDL* and *HDL*.

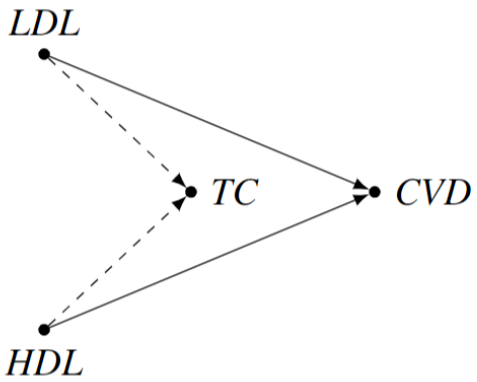


Figure 4.4 A "mixed" DAG containing both non-causal relations and causal relations.

However, the sort of mixed graph in Figure 4.4 is nothing like a standard DAG. It is widely acknowledged and observed in causal inference practice that variables in a standard DAG should not stand in non-causal relationships. Therefore, there is no standard DAG representation for the causal structure underlying the type of SP we are discussing here. In particular, *HDL* in Figure 4.4 is *not* a confounding variable since it is not even a cause of *TC*. This means that Pearl's causal-graphical analysis cannot handle this type of SP.

How should we explain away and avoid this type of paradox then? The answer is simple: *TC* should not be used as a cause of *CVD* to begin with, because its "effect" on *CVD* is ambiguous. Instead, we should use *HDL* and *LDL* as causes when investigating the incidence of cardiovascular diseases. So, it turns out that the genuine source of SP in this case is not confounding but a bad choice of causal variables, that is, the use of inappropriately summed variables as a cause.

4.3 Inter-Unit Causation

Following the broader point behind Section 0, we now present another type of SP that can arise due to inappropriate variable choice instead of confounding. This type of SP has its root in having chosen a set of variables which are defined on a group of units that causally interact with each other. This phenomenon is known as *inter-unit causation* (Spirtes et al., 2000, p. 296; Zhang & Spirtes, 2014) or interactions among units (Zhang et al., 2022). As we shall see, inter-unit causation may create an association between two causally independent variables, giving rise to SP. This type of SP is not due to confounding, and thus falls outside of Pearl's (2014) analysis; a more nuanced causal analysis is needed.

Such type of SP can be found, for example, in non-stationary time series data, a series of data whose statistical properties (e.g., mean) depend on the time when the data are collected.[39] Let us consider a concrete example (adapted from Hoover, 2003), which we call "Height&Math":

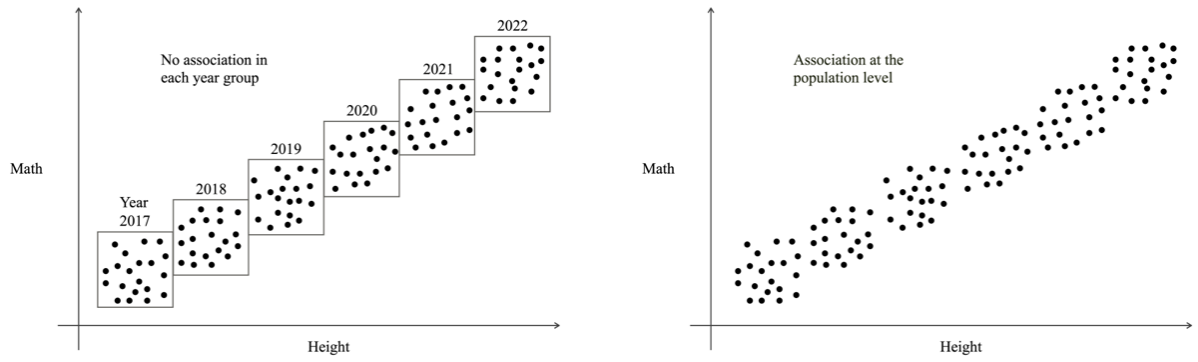
> **Height&Math:** Choose a class of 20 six-year-old children in the US and a class of 20 six-year-old children in China. Each year, we first measure the heights of the US children (*Height*) and order the data in the alphabetic order of their last names, and then measure

[39] Perhaps the most well-known example of this type is Sober's (2001) Venetian sea levels and British bread prices example (cf. Hoover, 2003; Steel, 2003). However, it is not easy to intuitively demonstrate why Sober's example is a case of SP: when we fix the year in which sea levels and bread prices are measured, the sample size reduces to 1, which makes it hard to show that the association between them disappears when conditioning on year.

the mathematical knowledge of the Chinese children (*Math*) using a standard diagnostic test, and order data in the same way.[40] Collect the data annually for six years.

In every single year, a US child's height and a Chinese child's mathematical knowledge are expected to be *unassociated*. The fact that a US child is of a certain height (e.g., 43 inches) should indicate nothing about what score the counterpart Chinese child earns on the math test, and vice versa. Yet, over the years, as the US children grow taller, the Chinese children also learn more math in school. As a result, in the data of all six years, both *Height* and *Math* will increase monotonically, producing a strong association between them. As illustrated in Figure 4.5, by treating the six years as a whole group and every single year as a sub-group, we see that an association absent at the sub-group level (Figure 4.5a) emerges at the group level (Figure 4.5b), making this example a case of SP.

[40] It is not a necessary feature of this example that the data be ordered by children's names. As long as we choose an ordering that is "random" and use it consistently throughout the data collection process, SP can arise.

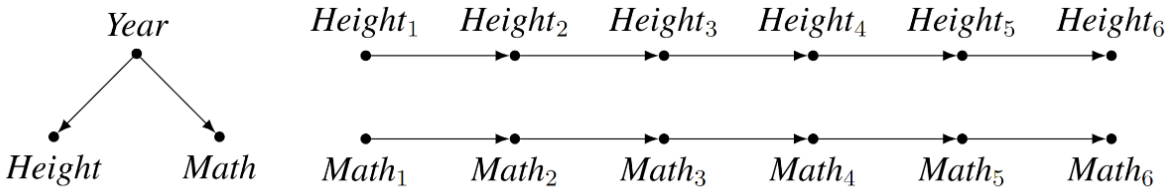


(a) When looking at data collected within each year (e.g., 2017), we observe no significant association

(b) When looking at the data collected over six years, we observe a strong association between *Math* and *Height*

Figure 4.5 An illustration of the Height&Math example.

Looking at this example, one might be tempted to posit *Year* as a common cause; doing so would allow this SP to be explained in terms of confounding. However, we believe this idea cannot stand scrutiny. Here, we agree with Yule (1926) and Steel (2003)—as well as Zhang and Spirtes (2014)—that it is ill-motivated to treat time as a cause. We are not saying this is uncontroversial, but given this broad consensus in the literature, the burden of proof is on those who think otherwise. The graph in Figure 4.6a is, therefore, not an appropriate representation of the causal structure behind the Height&Math example.



(a) A wrong DAG (b) A more accurate DAG with two series of variables

Figure 4.6 Two DAGs for the Height&Math example.

We believe the best diagnosis of how SP arises in this example is as follows (see Spirtes et al., 2000, p. 296; Zhang & Spirtes, 2014). So far, we have used "height" and "mathematical knowledge" as two ordinary random variables, which are measured over several years on the chosen children. However, a child's height (or math knowledge) in one year is *causally* relevant to the child's height (or math knowledge) in the next year, because the child in one year grows (or learns) based on the last year's height (or math knowledge). The original unit choice and variable choice thus induce "*inter-unit*" causation, meaning that there is a causal relationship between how one unit instantiates the relevant properties and how another unit instantiates them. In other words, the statistical units selected for analyzing a set of variables causally interact with each other. This creates trouble for standard statistical and causal analysis because inter-unit causation leads to the violation of the IID (independent and identically distributed) assumption in the sampling process. Causal analysis, built upon statistical analysis, is designed to capture *intra*-unit causation, not *inter*-unit causation (Zhang & Spirtes, 2014).

Given the above diagnosis, the solution is to redefine height and mathematical knowledge as two *series* of variables: $\mathbf{Height}_t = \{Height_1, ..., Height_6\}$ and $\mathbf{Math}_t = \{Math_1, ..., Math_6\}$, where each variable

is indexed to a year. This new unit choice and variable choice induces no inter-unit causation, and also successfully captures the underlying causal story. The resulting DAG is shown in Figure 4.6b, which represents increases in height and mathematical knowledge using two parallel but separate causal chains. Comparing Figure 4.6b against Figure 4.6a makes it clear that "[i]t is the inter-unit causation that propagates an initial coincidence into [an] association", as J. Zhang and Spirtes (2014, p. 247) summarize.

In our opinion, we need a DAG at least as complex as the one in Figure 4.6b to accurately depict the causal structure behind the kind of SP involved in the Height&Math example, and in general, any SP due to inter-unit causation.[41] Note this causal analysis has gone far beyond what Pearl's analysis can offer. Rather than simply drawing a DAG among the three given variables (i.e., *Height*, *Math* and *Year*) involved in the case of SP in the Height&Math example, what the new analysis demands is a redefinition of the variables and a reselection of units such that inter-unit causation is avoided.

[41] Of course, this more complicated DAG is still a *simplified* representation of the true causal structure. For example, a child's current math knowledge is influenced by not only her math knowledge in the previous year, but also the amount of education she received during the year. Representing these extra causes will make the DAG more complete and more accurate, but will not affect our conclusion that time should be understood as what the variables are indexed to, rather than a causal variable on its own.

5. The Amalgamation Paradox and the Non-Collapsibility of Odds Ratio

All cases of SP discussed in the previous two sections share the common feature that the X variable is associated with the partitioning variable Z (e.g., *Size* is associated with *Bag*; *Selfish* with *Group*, etc.). As mentioned earlier, the presence of this association is necessary for SP to occur. However, if we shift our focus from SP to its more generalized form, the *amalgamation paradox* (AP), the association between X and Z would no longer be necessary.[42]

AP is the statistical phenomenon in which the *marginal* or unconditional association of two variables, X and Y, falls outside the range of the *conditional* associations with respect to a third variable, Z. To illustrate this definition, consider an example from Greenland (2021). Suppose that a study on a medical treatment with high mortality has 50 males and 100 females in both the treatment and the control group. The two groups are relatively homogeneous in their medically relevant features such that no factor other than the treatment and sex can account for the difference in outcomes between the two groups. That is, in this example, *Treatment* is unassociated with *Sex*, and *Sex* is *not* a confounder (but an effect modifier; more below). Results show that among treated patients, 45 of 50 males and 30 of 100

[42] As noted by Samuels (1993, p. 84), AP was first identified and defined by Good and Mittal (1987) who pointed out that SP implies AP: all cases of SP are cases of AP. See also Sprenger and Weinberger (2021) for a definition of AP and its relationship with SP.

females died, whereas, among untreated patients, 30 of 50 males and 10 of 100 females died (see Table 4.4).

Table 4.4 This table demonstrates a case of AP relative to the odds ratio (OR) measure: the marginal association between *Treatment* and *Death* in the total group is 2.75, which is smaller than conditional associations in both males and females (6.0 and 3.9 respectively). For comparison, AP does not arise when risk difference is the association measure.

	Males		Females		Total	
	Treated	Untreated	Treated	Untreated	Treated	Untreated
Totals	50	50	100	100	150	150
Died	45	30	30	10	75	40
Risks/Proportions	0.90	0.60	0.30	0.10	0.500	0.267
Risk differences	0.30		0.20		0.233	
Odds	9/1	3/2	3/7	1/9	1/1	4/11
Odds ratios	**6.0**		**3.9**		**2.75**	

The above results indicate that the medical treatment has effects of *different magnitudes* on mortality among males and females. This is true regardless of whether the effect measure is odds ratio (OR) or risk difference—both are widely used in epidemiology. But there is something counterintuitive about OR in the above example. When OR is the chosen measure, we have OR(*Treatment*, *Death*) <

OR(*Treatment*, *Death* | *Sex* = Female) < OR(*Treatment*, *Death* | *Sex* = Male). This implies that the marginal association between *Treatment* and *Death* cannot be expressed as a weighted average of the two conditional associations with respect to *Sex*. That is, the marginal association falls outside the range of the conditional associations; this makes the example an instance of AP. Note that what gives rise to AP in the above example is the use of the odds ratio measure, which has been known to be *non-collapsible* (Hernán et al., 2011).[43] By contrast, no such peculiarity is present when the chosen measure is risk difference, which is a *collapsible* measure, meaning that on this measure, the marginal association falls between the conditional associations.[44]

More broadly speaking, the peculiarity of AP lies in the fact that it violates our expectation that the association between two variables at the level of the entire group should be bounded by their

[43] See also Cummings (2009) for more discussions on the non-collapsibility of odds ratios.

[44] The notion of (non-)collapsibility used here refers primarily to a property of an association measure. A measure, m, of the association between X and Y is collapsible across Z, if and only if the measured marginal association $m(X, Y)$ can be expressed as a *weighted average* of the measured conditional associations, $m(X, Y \mid Z)$ (cf. Pearl, 2009, p. 193; Huitfeldt et al., 2019). Risk difference, for instance, is a collapsible association measure in this sense. In contrast, odds ratio is a non-collapsible measure because there are datasets (such as the one represented in Table 4) where $OR(X, Y)$ cannot be expressed as a weighted average of $OR(X, Y \mid Z)$. One can also use collapsibility in a derivative sense that describe datasets. A dataset, D, is collapsible relative to a chosen association measure, m, if and only if the measured marginal association, $m(X, Y)$, reported in D can be expressed as a *weighted average* of the measured conditional associations, $m(X, Y \mid Z)$, in D. The dataset represented in Table 4 is non-collapsible relative to odds ratio in this sense. Moreover, Bandyopadhyay et al.'s (2015) notion of collapsibility states that a dataset is collapsible if and only if the marginal association between X and Y has the *same direction* as their conditional associations on Z. Since the marginal association being able to be expressed as a weighted average of the conditional associations implies that the marginal and the conditional associations are in the same direction, Bandyopadhyay et al.'s notion of (dataset) collapsibility encompasses the notion of (dataset) collapsibility as defined here in terms of weighted average. We thank an anonymous reviewer for prompting us to clarify this point.

conditional associations in the sub-groups (Hernán et al., 2011). This intuition comes from the seemingly plausible presupposition—which is not always true—that the marginal association should be a weighted average of the conditional associations. As a special case, SP violates a categorical or qualitative version of this intuition; namely, we expect that if two variables are conditionally unassociated or positively/negatively associated in each sub-group, they should likewise be unassociated or positively/negatively associated in the entire group. So, in cases of SP, not only the marginal association between X and Y falls outside the range of associations conditional on Z, but the marginal and the conditional associations are in *opposite* directions (e.g., the former is positive whereas the latter are non-positive). This is why SP, defined as association reversal, disappearance, or emergence, is a special form of AP. In light of this, it is natural to extend our attention to AP and ask the following questions: how can AP arise in a causal context, and when it does, what should we do? Attempts at answering these questions will help delineate the scope of the application of Pearl's causal-graphical analysis of SP.

What is particularly noteworthy is that the kind of AP due to a choice of non-collapsible association measure typically occurs in the presence of effect modification.[45] Effect modification occurs

[45] In other words, the combination of effect modification and non-collapsibility offers *an* important and typical causal context for the presence of AP in the data. However, this is not to say that effect modification is a necessary condition for the presence of AP. It is possible that the marginal odds ratio between X and Y differs from the conditional odds ratios when the latter two are equal. Nevertheless, the occurrence of AP without effect modification is much less frequent in real life, as well as less impressive, compared to cases in which AP is present due to effect modification.

when a causal effect is sensitive to the value of a third variable.[46] That is, Z is a *modifier* of the effect of X on Y when the effect varies across levels of Z (cf. Hernán & Robins, 2020). Importantly, a modifier is not a confounding variable. In our above example, *Sex* is a modifier of the treatment effect, but it is not a confounding variable since it is not associated with *Treatment*. This means that AP can occur in the absence of confounding. For this reason, Pearl's analysis of SP cannot be generalized to AP. To be fair, we are not claiming that this is an intrinsic difficulty with Pearl's analysis of SP, since to our knowledge, he does not claim that his analysis can be extended to dealing with AP. Still, we believe our discussion here helps us see the scope of Pearl's analysis.

Given that non-collapsibility is necessary for the type of AP discussed in this section, a general solution is to avoid using non-collapsible effect measures such as odds ratios without well-justified reasons. As for circumstances where we do want to use the odds ratio, we should be aware that compared with causal effect at the whole-group level, the effect of X on Y measured in the sub-groups stratified on the modifier Z conveys more accurate causal information about how X influences Y in a specific causal background. Thus, when the data display a pattern of AP due to non-collapsibility in the presence of effect modification, we should report sub-group-specific effect estimations (although this does not imply that effect estimation at the whole-group level is biased or meaningless).

[46] It is not easy to represent the presence of effect modification using standard DAGs; but see Weinberg (2007) for attempts of clarifying effect modification using non-standard DAGs.

6. Conclusion

In this chapter, we have argued that Pearl's causal-graphical analysis of SP, which diagnoses SP as essentially a peculiar consequence of confounding, cannot capture the full spectrum of the phenomenon. We show that there are good reasons to believe that SP is a generic term encompassing a wide range of distinct phenomena. Confounding is by no means the only source of SP, even if we admit that it is probably the most common and important one. Importantly, we do not claim that we have identified all possible sources of SP. There is nothing surprising about this, given that spurious associations observed in statistical data may come from a variety of sources: confounding, inappropriate variable aggregation, inter-unit causation, or sheer chance. The multiplicity of the sources of spurious association necessitates the multiplicity of the sources of SP. Thus, contra Pearl et al. (2016), we find it untenable that a confounding-based analysis can resolve all cases of SP. As far as we can see, the plurality of the sources of SP, and thereby the plurality of its resolutions, are here to stay.

Acknowledgements

Chapter 4, in full, is a reprint of the paper co-authored with Zili Dong and Shimin Zhao, as it appears in *European Journal for Philosophy of Science*, 2024. The dissertation author was the corresponding and co-first author of this paper.

References

Abbott, R. D., P. W. F. Wilson, W. B. Kannel, and W. P. Castelli (1988). "High Density Lipoprotein Cholesterol, Total Cholesterol Screening, and Myocardial Infarction: The Framingham Study." *Arteriosclerosis* 8 (3): 207–211.

Adams, F., and K. Aizawa (2021). "Causal Theories of Mental Content." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. Metaphysics Research Lab, Stanford University.

Adjodah, D., K. Dinakar, M. Chinazzi, S. P. Fraiberger, A. Pentland, S. Bates, K. Staller, A. Vespignani, and D. L. Bhatt (2021). "Association Between COVID-19 Outcomes and Mask Mandates, Adherence, and Attitudes." *PLoS ONE* 16 (6): e0252315.

Andreas, H., and M. Günther (2021). "Difference-Making Causation." *Journal of Philosophy* 118 (12): 680–701.

Antecol, H. (2000). "An Examination of Cross-Country Differences in the Gender Gap in Labor Force Participation Rates." *Labour Economics* 7 (4): 409–426.

Anscombe, G. E. M. (1993). "Causality and Determination." In *Causation*, edited by E. Sosa and M. Tooley, 88–104. Oxford University Press.

Bandyopadhyay, P. S., M. Greenwood, D. W. F. Dcruz, and V. Raghavan R. (2015). "Simpson's Paradox and Causality." *American Philosophical Quarterly* 52 (1): 13–25.

Bandyopadhyay, P. S., D. Nelson, M. Greenwood, G. Brittan, and J. Berwald (2011). "The Logic of Simpson's Paradox." *Synthese* 181 (2): 185–208.

Baumgartner, M. (2008). "The Causal Chain Problem." *Erkenntnis* 69 (2): 201–226.

Baumgartner, M. (2013). "A Regularity Theoretic Approach to Actual Causation." *Erkenntnis* 78 (Suppl. 1): 85–109.

Baumgartner, M., and A. Gebharter (2016). "Constitutive Relevance, Mutual Manipulability, and Fat-Handedness." *British Journal for the Philosophy of Science* 67 (3): 731–756.

Baumgartner, M., and C. Falk (2023). "Boolean Difference-Making: A Modern Regularity Theory of Causation." *British Journal for the Philosophy of Science* 74 (1): 171–197.

Bebb, J., and H. Beebee (2024). "Causal Selection and Egalitarianism." In *Oxford Studies in Experimental Philosophy*, Volume 5, edited by S. Nichols and J. Knobe, 401–433. Oxford University Press.

Bechtel, W., and A. Abrahamsen (2005). "Explanation: A Mechanist Alternative." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 36 (2): 421–441.

Bechtel, W., and A. Bollhagen (2024). "Causal Cycles in Biology." In *The Routledge Handbook of Causality and Causal Methods*, edited by P. Illari and F. Russo, 255–268. Routledge.

Beebee, H., C. Hitchcock, and H. Price, eds. (2017). *Making a Difference: Essays on the Philosophy of Causation*. Oxford University Press.

Bell, J. S. (2004). *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy*. 2nd ed. Cambridge University Press.

Bennett, J. (1988). *Events and their Names*. Hackett Publishing Company.

Berger, S., ed. (2009). *The Foundations of Non-Equilibrium Economics: The Principle of Circular and Cumulative Causation*. Routledge.

Bickel, P. J., E. A. Hammel, and J. W. O'Connell (1975). "Sex Bias in Graduate Admissions: Data from Berkeley." *Science* 187 (4175): 398–404.

Bird, A. (2020). "A Dispositional Account of Causation, with Some Remarks on the Ontology of Dispositions." In *Dispositionalism: Perspectives from Metaphysics and the Philosophy of Science*, edited by A. S. Meincke, 151–170. Springer.

Broadbent, A. (2008). "The Difference between Cause and Condition." *Proceedings of the Aristotelian Society* 108 (1): 355–364.

Broadbent, A. (2015). "Causation and Prediction in Epidemiology: A Guide to the 'Methodological Revolution'." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 54: 72–80.

Brown, M. L., A. Inazu, C. B. Hesler, L. B. Agellon, C. Mann, M. E. Whitlock, Y. L. Marcel, R. W. Milne, J. Koizumi, H. Mabuchi, R. Takeda, and A. R. Tall (1989). "Molecular Basis of Lipid Transfer Protein Deficiency in a Family with Increased High-Density Lipoproteins." *Nature* 342 (6248): 448–451.

Cartwright, N. (1979). "Causal Laws and Effective Strategies." *Noûs* 13 (4): 419–437.

Cartwright, N. (1989). *Nature's Capacities and Their Measurement*. Clarendon Press.

Cartwright, N. (1999). "Causal Diversity and the Markov Condition." *Synthese* 121 (1/2): 3–27.

Cartwright, N. (2001). "What Is Wrong with Bayes Nets?" *The Monist* 84 (2): 242–264.

Cartwright, N. (2004). "Causation: One Word, Many Things." *Philosophy of Science* 71 (5): 805–819.

Cartwright, N. (2006). "Well-Ordered Science: Evidence for Use." *Philosophy of Science* 73 (5): 981–990.

Cartwright, N. (2020). "Middle-Range Theory: Without It What Could Anyone Do?" *Theoria* 35 (3): 269–323.

Cartwright, N. (2021). "Rigour Versus the Need for Evidential Diversity." *Synthese* 199 (5–6): 13095–13119.

Cartwright, N., and J. Hardie (2012). *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford University Press.

Cartwright, N., B. Kelters, E. Munro, L. Ncube, and J. Pemberton (2026). "Can I Make This Policy Work Here? Better Theories of Change and Better Evidence Management for Better Policy Design." CHESS Working Paper No. 2026-01, Durham University

Castelli, W. P., J. T. Doyle, T. Gordon, C. G. Hames, M. C. Hjortland, S. B. Hulley, A. Kagan, and W. J. Zukel (1977). "HDL Cholesterol and Other Lipids in Coronary Heart Disease: The Cooperative Lipoprotein Phenotyping Study." *Circulation* 55 (5): 767–772.

Caudill, W. A. (1973). "The Influence of Social Structure and Culture on Human Behavior in Modern Japan." *The Journal of Nervous and Mental Disease* 157 (4): 240–257.

Cheng, P. W., and L. R. Novick (1991). "Causes Versus Enabling Conditions." *Cognition* 40 (1–2): 83–120.

Child, W. (2020). "Economics, Agency, and Causal Explanation." In *Agency and Causal Explanation in Economics*, edited by P. Róna and L. Zsolnai, 53–67. Springer.

Cicchetti, D., and F. A. Rogosch (1996). "Equifinality and Multifinality in Developmental Psychopathology." *Development and Psychopathology* 8 (4): 597–600.

Claveau, F. (2012). "The Russo–Williamson Theses in the Social Sciences: Causal Inference Drawing on Two Types of Evidence." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (4): 806–813.

Costa, F., and S. Shrapnel (2016). "Quantum Causal Modelling." *New Journal of Physics* 18 (6): Article 063032.

Craver, C. F. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Clarendon Press.

Craver, C. F., and L. Darden (2013). *In Search of Mechanisms: Discoveries across the Life Sciences*. University of Chicago Press.

Craver, C. F., S. Glennan, and M. Povich (2021). "Constitutive Relevance & Mutual Manipulability Revisited." *Synthese* 199 (3–4): 8807–8828.

Cummings, P. (2009). "The Relative Merits of Risk Ratios and Odds Ratios." *Archives of Pediatrics & Adolescent Medicine* 163 (5): 438–445.

Danks, D. (2013). "Functions and Cognitive Bases for the Concept of Actual Causation." *Erkenntnis* 78 (Suppl. 1): 111–128.

David, P. A. (1985). "Clio and the Economics of QWERTY." *American Economic Review* 75 (2): 332–337.

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

Dawid, A. P. (2010). "Beware of the DAG!" In *Proceedings of Workshop on Causality: Objectives and Assessment at NIPS 2008*, *Proceedings of Machine Learning Research* 6: 59–86.

Didelez, V., and M. J. Stensrud (2022). "On the Logic of Collapsibility for Causal Effect Measures." *Biometrical Journal* 64 (2): 235–242.

Dobzhansky, T., and O. Pavlovsky (1957). "An Experimental Study of Interaction between Genetic Drift and Natural Selection." *Evolution* 11 (3): 311–319.

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

Dretske, F. I. (1981). *Knowledge and the Flow of Information*. MIT Press.

Dummett, M. (1964). "Bringing About the Past." *The Philosophical Review* 73 (3): 338–359.

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

Egami, N., and K. Imai (2019). "Causal Interaction in Factorial Experiments: Application to Conjoint Analysis." *Journal of the American Statistical Association* 114 (526): 529–540.

Ellis, B. (2001). *Scientific Essentialism*. Cambridge University Press.

Fair, D. (1979). "Causation and the Flow of Energy." *Erkenntnis* 14 (3): 219–250.

Falleti, T. G., and J. F. Lynch (2009). "Context and Causal Mechanisms in Political Analysis." *Comparative Political Studies* 42 (9): 1143–1166.

Feder, H. M., Jr., B. W. Petersen, K. L. Robertson, and C. E. Rupprecht (2012). "Rabies: Still a Uniformly Fatal Disease? Historical Occurrence, Epidemiological Trends, and Paradigm Shifts." *Current Infectious Disease Reports* 14 (4): 408–422.

Fernandez, V. (2014). "Linear and Non-Linear Causality between Price Indices and Commodity Prices." *Resources Policy* 41: 40–51.

Flaherty, M. L., H. Tao, M. Haverbusch, P. Sekar, D. Kleindorfer, B. Kissela, P. Khatri, B. Stettler, O. Adeoye, C. J. Moomaw, J. P. Broderick, and D. Woo (2008). "Warfarin Use Leads to Larger Intracerebral Hematomas." *Neurology* 71 (14): 1084–1089.

Frisch, M. (2020). "Causation in Physics." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta and U. Nodelman, Metaphysics Research Lab, Stanford University.

Gallus, C., E. M. Pothos, P. Blasiak, J. M. Yearsley, and B. W. Wojciechowski (2023). "Bell Correlations Outside Physics." *Scientific Reports* 13: 4394.

Ganson, T. (2021). "An Alternative to the Causal Theory of Perception." *Australasian Journal of Philosophy* 99 (4): 683–695.

Gerring, J. (2011). *Social Science Methodology: A Unified Framework*. 2nd ed. Cambridge University Press.

Ghiara, V. (2022). "Taking the Russo–Williamson Thesis Seriously in the Social Sciences." *Synthese* 200 (6): Article 481.

Glennan, S. S. (1996). "Mechanisms and the Nature of Causation." *Erkenntnis* 44 (1): 49–71.

Glennan, S. S. (2017). *The New Mechanical Philosophy*. Oxford University Press.

Glymour, C., D. Danks, B. Glymour, F. Eberhardt, J. Ramsey, R. Scheines, P. Spirtes, C. M. Teng, and J. Zhang (2010). "Actual Causation: A Stone Soup Essay." *Synthese* 175 (2): 169–192.

Glynn, L. (2011). "A Probabilistic Analysis of Causation." *British Journal for the Philosophy of Science* 62 (2): 343–392.

Goldman, A. I. (1967). "A Causal Theory of Knowing." *Journal of Philosophy* 64 (12): 357–372.

Good, I. J., and Y. Mittal (1987). "The Amalgamation and Geometry of Two-by-Two Contingency Tables." *The Annals of Statistics* 15 (2): 694–711.

Goodwin, W. (2012). "Mechanisms and Chemical Reaction." In *Philosophy of Chemistry*, edited by A. I. Woody, R. F. Hendry, and P. Needham, 309–327. North-Holland.

Gordon, D. J., J. L. Probstfield, R. J. Garrison, J. D. Neaton, W. P. Castelli, J. D. Knoke, D. R. Jacobs Jr., S. Bangdiwala, and H. A. Tyroler (1989). "High-Density Lipoprotein Cholesterol and Cardiovascular Disease: Four Prospective American Studies." *Circulation* 79 (1): 8–15.

Gourévitch, B., R. Le Bouquin-Jeannès, and G. Faucon (2006). "Linear and Nonlinear Causality between Signals: Methods, Examples and Neurophysiological Applications." *Biological Cybernetics* 95 (4): 349–369.

Green, S., and N. J. Jones (2016). "Constraint-Based Reasoning for Search and Explanation: Strategies for Understanding Variation and Patterns in Biology." *Dialectica* 70 (3): 343–374.

Greenland, S. (2021). "Noncollapsibility, Confounding, and Sparse-Data Bias. Part 2: What Should Researchers Make of Persistent Controversies about the Odds Ratio?" *Journal of Clinical Epidemiology* 139: 264–268.

Grice, H. P. (1961). "The Causal Theory of Perception." *Proceedings of the Aristotelian Society, Supplementary Volumes* 35 (1): 121–152.

Gross, N. (2018). "The Structure of Causal Chains." *Sociological Theory* 36 (4): 343–367.

Hall, N. (2000). "Causation and the Price of Transitivity." *Journal of Philosophy* 97 (4): 198–222.

Hall, N. (2004). "Two Concepts of Causation." In *Causation and Counterfactuals*, edited by J. Collins, N. Hall, and L. A. Paul, 225–276. MIT Press.

Hargens, L. L. (2009). "Product-Variable Models of Interaction Effects and Causal Mechanisms." *Social Science Research* 38 (1): 19–28.

Harré, R. (2016). "Causality in Chemistry: Regularities and Agencies." In *Essays in the Philosophy of Chemistry*, edited by E. Scerri and G. Fisher, 179–198. Oxford University Press.

Hart, H. L. A., and T. Honoré (1985). *Causation in the Law*. 2nd ed. Oxford University Press.

Hedström, P. (2005). *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge University Press.

Hedström, P., and L. Udehn (2009). "Analytical Sociology and Theories of the Middle Range." In *The Oxford Handbook of Analytical Sociology*, edited by P. Hedström and P. Bearman, 25–50. Oxford University Press.

Hedström, P., and P. Ylikoski (2010). "Causal Mechanisms in the Social Sciences." *Annual Review of Sociology* 36 (1): 49–67.

Hendry, R. F. (2023). "Mechanisms in Chemistry." In *New Mechanism Explanation, Emergence and Reduction*, edited by J. L. Cordovil, G. Santos, and D. Vecchi, 139–160. Springer.

Hernán, M. A., D. Clayton, and N. Keiding (2011). "The Simpson's Paradox Unraveled." *International Journal of Epidemiology* 40 (3): 780–785.

Hernán, M. A., and J. M. Robins (2020). *Causal Inference: What If*. Chapman & Hall/CRC.

Hernández-Díaz, S., E. F. Schisterman, and M. A. Hernán (2006). "The Birth Weight 'Paradox' Uncovered?" *American Journal of Epidemiology* 164 (11): 1115–1120.

Hesslow, G. (1976). "Two Notes on the Probabilistic Approach to Causality." *Philosophy of Science* 43 (2): 290–292.

Hitchcock, C. (1995). "The Mishap at Reichenbach Fall: Singular vs. General Causation." *Philosophical Studies* 78 (3): 257–291.

Hitchcock, C. (1999). "Contrastive Explanation and the Demons of Determinism." *British Journal for the Philosophy of Science* 50 (4): 585–612.

Hitchcock, C. (2001). "The Intransitivity of Causation Revealed in Equations and Graphs." *Journal of Philosophy* 98 (6): 273–299.

Hoover, K. D. (2001). *Causality in Macroeconomics*. Cambridge University Press.

Hoover, K. D. (2003). "Nonstationary Time Series, Cointegration, and the Principle of the Common Cause." *British Journal for the Philosophy of Science* 54 (4): 527–551.

Hsiang, S., D. Allen, S. Annan-Phan, K. Bell, I. Bolliger, T. Chong, H. Druckenmiller, L. Y. Huang, A. Hultgren, E. Krasovich, P. Lau, J. Lee, E. Rolf, J. Tseng, and T. Wu (2020). "The Effect of Large-Scale Anti-Contagion Policies on the COVID-19 Pandemic." *Nature* 584 (7820): 262–267.

Huitfeldt, A., M. J. Stensrud, and E. Suzuki (2019). "On the Collapsibility of Measures of Effect in the Counterfactual Causal Framework." *Emerging Themes in Epidemiology* 16: 1–5.

Hume, D. (1739/1986). *A Treatise of Human Nature*. Penguin Classics.

Illari, P. M. (2011). "Mechanistic Evidence: Disambiguating the Russo–Williamson Thesis." *International Studies in the Philosophy of Science* 25 (2): 139–157.

Illari, P. M., and J. Williamson (2012). "What Is a Mechanism? Thinking about Mechanisms across the Sciences." *European Journal for Philosophy of Science* 2 (1): 119–135.

Ioannidis, S., and S. Psillos (2022). *Mechanisms in Science: Method or Metaphysics?* Cambridge University Press.

Irzik, G., and E. Meyer (1987). "Causal Modeling: New Directions for Statistical Explanation." *Philosophy of Science* 54 (4): 495–514.

Jacobs, A. M. (2014). "Process Tracing the Effects of Ideas." In *Process Tracing: From Metaphor to Analytic Tool*, edited by A. Bennett and J. T. Checkel, 41–73. Cambridge University Press.

Jansen, H., and W. C. Hülsmann (1980). "Heparin-Releasable (Liver) Lipase(s) May Play a Role in the Uptake of Cholesterol by Steroid-Secreting Tissues." *Trends in Biochemical Sciences* 5 (10): 265–268.

Jauk, E., M. Benedek, B. Dunst, and A. C. Neubauer (2013). "The Relationship between Intelligence and Creativity: New Support for the Threshold Hypothesis by Means of Empirical Breakpoint Detection." *Intelligence* 41 (4): 212–221.

Karbownik, K., and A. Wray (2019). "Long-Run Consequences of Exposure to Natural Disasters." *Journal of Labor Economics* 37 (3): 949–1007.

Kim, J. (1973). "Causation, Nomic Subsumption, and the Concept of Event." *Journal of Philosophy* 70 (8): 217–236.

Kim, J. (1976). "Events as Property Exemplifications." In *Action Theory*, edited by M. Brand and D. Walton, 310–326. Reidel.

Kim, J. (2005). *Physicalism, or Something Near Enough*. Princeton University Press.

Laland, K. N., K. Sterelny, J. Odling-Smee, W. Hoppitt, and T. Uller (2011). "Cause and Effect in Biology Revisited: Is Mayr's Proximate–Ultimate Dichotomy Still Useful?" *Science* 334 (6062): 1512–1516.

Lange, M. (2013). "What Makes a Scientific Explanation Distinctively Mathematical?" *British Journal for Philosophy of Science* 64 (3): 485–511.

Lange, M. (2023). "Explanations by Constraint: Not Just in Physics." *International Studies in the Philosophy of Science* 36 (4): 265–277.

Lewis, D. K. (1973). "Causation." *Journal of Philosophy* 70 (17): 556–567.

Lewis, D. K. (1976). "The Paradoxes of Time Travel." *American Philosophical Quarterly* 13 (2): 145–152.

Lewis, D. K. (1986a). "Events." In *Philosophical Papers, Volume II*, 241–269. Oxford University Press.

Lewis, D. K. (1986b). "Causal Explanation." In *Philosophical Papers, Volume II*, 214–240. Oxford University Press.

Lewis, D. K. (1986c). "Postscripts to 'Causation'." In *Philosophical Papers: Volume II*, 172–213. Oxford University Press.

Lewis, D. K. (2000). "Causation as Influence." *Journal of Philosophy* 97 (4): 182–197.

Li, Z.-H., Y.-B. Lv, W.-F. Zhong, X. Gao, V. Byers Kraus, M.-C. Zou, X.-R. Zhang, F.-R. Li, J.-Q. Yuan, X.-M. Shi, X.-B. Wu, and C. Mao (2019). "High-Density Lipoprotein Cholesterol and All-Cause and Cause-Specific Mortality among the Elderly." *Journal of Clinical Endocrinology & Metabolism* 104 (8): 3370–3378.

Lipton, P. (1990). "Contrastive Explanation." *Royal Institute of Philosophy Supplement* 27: 247–266.

Lipton, P. (2009). "Causation and Explanation." In *The Oxford Handbook of Causation*, edited by H. Beebee, C. Hitchcock, and P. Menzies, 619–631. Oxford University Press.

List, C., and P. Menzies (2009). "Nonreductive Physicalism and the Limits of the Exclusion Principle." *Journal of Philosophy* 106 (9): 475–502.

Liu, C., D. Dhindsa, Z. Almuwaqqat, Y. KO, A. Mehta, A. A. Alkhoder, Z. Alras, S. R. Desai, K. J. Patel, A. Hooda, M. Wehbe, L. S. Sperling, Y. V. Sun, and A. A. Quyyumi (2022a). "Association between High-Density Lipoprotein Cholesterol Levels and Adverse Cardiovascular Outcomes in High-Risk Populations." *JAMA Cardiology* 7 (7): 672–680.

Liu, C., D. Dhindsa, Z. Almuwaqqat, Y. V. Sun, and A. A. Quyyumi (2022b). "Very High High-Density Lipoprotein Cholesterol Levels and Cardiovascular Mortality." *American Journal of Cardiology* 167: 43–53.

Loew, C. (2019). "Causes as Difference-Makers for Processes." *Philosophy and Phenomenological Research* 98 (1): 89–106.

Lombard, L. B. (1990). "Causes, Enablers, and the Counterfactual Analysis." *Philosophical Studies* 59 (2): 195–211.

Lowe, E. J. (1980). "For Want of a Nail." *Analysis* 40 (1): 50–52.

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

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

Mackie, P. (1992). "Causing, Delaying, and Hastening: Do Rains Cause Fires?" *Mind* 101 (403): 483–500.

Madsen, C. M., A. Varbo, and B. G. Nordestgaard (2017). "Extreme High High-Density Lipoprotein Cholesterol Is Paradoxically Associated with High Mortality in Men and Women: Two Prospective Cohort Studies." *European Heart Journal* 38 (32): 2478–2486.

Mayr, E. (1961). "Cause and Effect in Biology: Kinds of Causes, Predictability, and Teleology Are Viewed by a Practicing Biologist." *Science* 134 (3489): 1501–1506.

Maziarz, M. (2020). *The Philosophy of Causality in Economics: Causal Inferences and Policy Proposals*. Routledge.

McDonnell, N. (2018). "Transitivity and Proportionality in Causation." *Synthese* 195 (3): 1211–1229.

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

Menzies, P. (1989). "A Unified Account of Causal Relata." *Australasian Journal of Philosophy* 67 (1): 59–83.

Menzies, P. (2004). "Difference-Making in Context." In *Causation and Counterfactuals*, edited by J. Collins, N. Hall, and L. A. Paul, 139–180. MIT Press.

Menzies, P., and H. Price (1993). "Causation as a Secondary Quality." *British Journal for the Philosophy of Science* 44 (2): 187–203.

Merton, R. K. (1968). "On Sociological Theories of the Middle Range." In *Social Theory and Social Structure*, enlarged ed., 39–72. The Free Press. Original work published 1949.

Mill, J. S. (1874). *A System of Logic, Ratiocinative and Inductive*. 8th ed. Harper & Brothers.

Molnar, G. (2003). *Powers: A Study in Metaphysics*. Clarendon Press.

Mukherjee, S. (2015). *The Laws of Medicine: Field Notes from an Uncertain Science*. New York: Simon & Schuster/TED.

Mulaik, S. A. (2009). *Linear Causal Modeling with Structural Equations*. Chapman and Hall/CRC.

Mumford, S., and R. L. Anjum (2011). *Getting Causes from Powers*. Oxford University Press.

Myrvold, W., M. Genovese, and A. Shimony (2024). "Bell's Theorem." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta and U. Nodelman, Metaphysics Research Lab, Stanford University.

Näger, P. M. (2022). "Evidence for Interactive Common Causes. Resuming the Cartwright-Hausman-Woodward Debate." *European Journal for Philosophy of Science* 12, 2.

Nolen-Hoeksema, S., and E. R. Watkins (2011). "A Heuristic for Developing Transdiagnostic Models of Psychopathology: Explaining Multifinality and Divergent Trajectories." *Perspectives on Psychological Science* 6 (6): 589–609.

Northcott, R. (2008). "Causation and Contrast Classes." *Philosophical Studies* 139 (1): 111–123.

Norton, J. D. (2003). "Causation as Folk Science." *Philosophers' Imprint* 3 (4): 1–22.

Norton, J. D. (2024). "The Metaphysics of Causation: An Empiricist Critique." In *Alternative Approaches to Causation: Beyond Difference-Making and Mechanism*, edited by Y. Shan, 58–94. Oxford University Press.

Paul, L. A. (2000). "Aspect Causation." *Journal of Philosophy* 97 (4): 235–256.

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

Pearl, J. (2014). "Comment: Understanding Simpson's Paradox." *The American Statistician* 68 (1): 8–13.

Pearl, J., M. Glymour, and N. P. Jewell (2016). *Causal Inference in Statistics: A Primer*. John Wiley & Sons.

Pearl, J., and D. Mackenzie (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books.

Pierson, P. (2000a). "Increasing Returns, Path Dependence, and the Study of Politics." *American Political Science Review* 94 (2): 251–267.

Pierson, P. (2000b). "Not Just What, but When: Timing and Sequence in Political Processes." *Studies in American Political Development* 14 (1): 72–92.

Pincock, C. (2018). "Accommodating Explanatory Pluralism." In *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*, edited by A. Reutlinger and J. Saatsi, 39–56. Oxford University Press.

Price, H. (1996). *Time's Arrow and Archimedes' Point: New Directions for the Physics of Time*. Oxford University Press.

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

Reiss, J. (2015). *Causation, Evidence, and Inference*. Routledge.

Ross, L. N. (2018). "Causal Selection and the Pathway Concept." *Philosophy of Science* 85 (4): 551–572.

Ross, L. N. (2025). *Explanation in Biology*. Cambridge University Press.

Ross, L. N. (2026). "Cascade versus Mechanism: The Diversity of Causal Structure in Science." *British Journal for the Philosophy of Science* 77 (1): 145–169.

Ross, L. N. (manuscript). "Causal Pluralism: Primary and Secondary Features of Causation."

Ross, L. N., and J. F. Woodward (2022). "Irreversible (One-Hit) and Reversible (Sustaining) Causation." *Philosophy of Science* 89 (5): 889–898.

Ross, L. N., and J. F. Woodward (2023). "Causal Approaches to Scientific Explanation." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta and U. Nodelman. Metaphysics Research Lab, Stanford University.

Rothblat, G. H., and M. C. Phillips (1982). "Mechanism of Cholesterol Efflux from Cells: Effects of Acceptor Structure and Concentration." *Journal of Biological Chemistry* 257 (9): 4775–4782.

Rupert, R. (2008). "Causal Theories of Mental Content." *Philosophy Compass* 3 (2): 353–380.

Russo, F., and J. Williamson (2007). "Interpreting Causality in the Health Sciences." *International Studies in the Philosophy of Science* 21 (2): 157–170.

Samuels, M. L. (1993). "Simpson's Paradox and Related Phenomena." *Journal of the American Statistical Association* 88 (421): 81–88.

Salmon, W. C. (1971). "Statistical Explanation." In *Statistical Explanation and Statistical Relevance*, 29–88. University of Pittsburgh Press.

Salmon, W. C. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton University Press.

Sarkar, P., and P. S. Bandyopadhyay (2021). "Simpson's Paradox: A Singularity of Statistical and Inductive Inference." arXiv:2103.16860.

Sartorio, C. (2005). "Causes as Difference-Makers." *Philosophical Studies* 123 (1/2): 71–96.

Sartorio, C. (2006). "On Causing Something to Happen in a Certain Way without Causing It to Happen." *Philosophical Studies* 129 (1): 119–136.

Sartorio, C. (2013). "Making a Difference in a Deterministic World." *The Philosophical Review* 122 (2): 189–214.

Schaffer, J. (2001). "Causes as Probability Raisers of Processes." *Journal of Philosophy* 98 (2): 75–92.

Schaffer, J. (2005). "Contrastive Causation." *Philosophical Review* 114 (3): 327–358.

Seifert, V. A. (2024). *Chemistry's Metaphysics*. Cambridge University Press.

Shan, Y. (2025). "The Debate over Proximate and Ultimate Causation in Biology." *Synthese* 205 (1): Article 37.

Shan, Y., and J. Williamson (2021). "Applying Evidential Pluralism to the Social Sciences." *European Journal for Philosophy of Science* 11 (4): Article 96.

Shan, Y., and J. Williamson (2022). "Evidential Monism, Evidential Pluralism, or Evidential Contextualism? An Introduction to Evidential Diversity in the Social Sciences." *Synthese* 200 (4): Article 321.

Shan, Y., and J. Williamson (2023). *Evidential Pluralism in the Social Sciences*. Routledge.

Shrier, I., and R. W. Platt (2008). "Reducing Bias through Directed Acyclic Graphs." *BMC Medical Research Methodology* 8 (1): 1–15.

Skocpol, T. (1979). *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. Cambridge University Press.

Skow, B. (2023a). "The Varieties of Relevance." In *Oxford Studies in Metaphysics: Volume 13*, edited by K. Bennett and D. W. Zimmerman, 71–128. Oxford University Press.

Skow, B. (2023b). "Causal Explanation." In *Routledge Encyclopedia of Philosophy*. Routledge.

Sober, E. (1984). "Two Concepts of Cause." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1984 (2): 405–424.

Sober, E. (2001). "Venetian Sea Levels, British Bread Prices, and the Principle of the Common Cause." *British Journal for the Philosophy of Science* 52 (2): 331–346.

Sober, E. (2020). "Contrastive Causal Explanation and the Explanatoriness of Deterministic and Probabilistic Hypotheses." *European Journal for Philosophy of Science* 10 (3): 1–15.

Sober, E. (2024). *The Philosophy of Evolutionary Theory: Concepts, Inferences, and Probabilities*. Cambridge University Press.

Spanos, A. (2021). "Yule–Simpson's Paradox: The Probabilistic versus the Empirical Conundrum." *Statistical Methods & Applications* 30 (2): 605–635.

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

Spirtes, P., and R. Scheines (2004). "Causal Inference of Ambiguous Manipulations." *Philosophy of Science* 71 (5): 833–845.

Sprenger, J., and N. Weinberger (2021). "Simpson's Paradox." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. Metaphysics Research Lab, Stanford University.

Stampe, D. (1977). "Toward a Causal Theory of Linguistic Representation." *Midwest Studies in Philosophy* 2 (1): 42–63.

Stanley, E. A., and J. P. Sawyer (2009). "The Equifinality of War Termination: Multiple Paths to Ending War." *Journal of Conflict Resolution* 53 (5): 651–676.

Steel, D. (2003). "Making Time Stand Still: A Response to Sober's Counter-Example to the Principle of the Common Cause." *British Journal for the Philosophy of Science* 54 (2): 309–317.

Stegmann, U. E. (2014). "Causal Control and Genetic Causation." *Noûs* 48 (3): 450–465.

Stephan, S. (2025). "Reasoning about Actual Causation in Reversible and Irreversible Causal Structures." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 51 (1): 152–169.

Strand, A., and G. Oftedal (2019). "Restricted Causal Relevance." *British Journal for the Philosophy of Science* 70 (2): 431–457.

Suárez, M. (2014). "The Contextual Character of Causal Evidence." *Topoi* 33 (2): 397–406.

Suppes, P. (1970). *A Probabilistic Theory of Causality*. North-Holland Publishing Company.

Tall, A. R. (1986). "Plasma Lipid Transfer Proteins." *Journal of Lipid Research* 27 (4): 361–367.

Touborg, C. T. (2018). "Hasteners and Delayers: Why Rains Don't Cause Fires." *Philosophical Studies* 175 (7): 1557–1576.

Tran, B. L., C. C. Chen, and W. C. Tseng (2022). "Causality between Energy Consumption and Economic Growth in the Presence of GDP Threshold Effect: Evidence from OECD Countries." *Energy* 251: Article 123902.

van Fraassen, B. (1980). *The Scientific Image*. The Clarendon Press.

VanderWeele, T. J., and J. M. Robins (2009). "Properties of Monotonic Effects on Directed Acyclic Graphs." *Journal of Machine Learning Research* 10 (24): 699–718.

Veltkamp, R., T. Rizos, and S. Horstmann (2013). "Intracerebral Bleeding in Patients on Antithrombotic Agents." *Seminars in Thrombosis and Hemostasis* 39 (8): 963–971.

Warrell, J., and M. Gerstein (2020). "Cyclic and Multilevel Causation in Evolutionary Processes." *Biology & Philosophy* 35: Article 50.

Waters, C. K. (2007). "Causes That Make a Difference." *Journal of Philosophy* 104 (11): 551–579.

Weinberg, C. R. (2007). "Can DAGs Clarify Effect Modification?" *Epidemiology* 18 (5): 569–572.

Weinberger, N., P. Williams, and J. Woodward (forthcoming). "The Worldly Infrastructure of Causation." *British Journal for the Philosophy of Science*.

Wheeler, G., and R. Scheines (2013). "Coherence and Confirmation through Causation." *Mind* 122 (485): 135–170.

Wilde, M., and V.-P. Parkkinen (2019). "Extrapolation and the Russo–Williamson Thesis." *Synthese* 196 (8): 3251–3262.

Williamson, J. (2005). *Bayesian Nets and Causality: Philosophical and Computational Foundations*. Oxford University Press.

Williamson, J. (2006a). "Causal Pluralism versus Epistemic Causality." *Philosophica* 77 (1): 69–96.

Williamson, J. (2006b). "Dispositional versus Epistemic Causality." *Minds and Machines* 16 (3): 259–276.

Wood, C. J., and R. W. Spekkens (2015). "The Lesson of Causal Discovery Algorithms for Quantum Correlations: Causal Explanations of Bell-Inequality Violations Require Fine-Tuning." *New Journal of Physics* 17 (3): 033002.

Woodward, J. (2002). "What Is a Mechanism? A Counterfactual Account." *Philosophy of Science* 69 (S3): S366–S377.

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

Woodward, J. (2010). "Causation in Biology: Stability, Specificity, and the Choice of Levels of Explanation." *Biology & Philosophy* 25 (3): 287–318.

Woodward, J. (2014). "A Functional Account of Causation; or, A Defense of the Legitimacy of Causal Thinking by Reference to the Only Standard That Matters—Usefulness." *Philosophy of Science* 81 (5): 691–713.

Woodward, J. (2015a). "Methodology, Ontology, and Interventionism." *Synthese* 192 (11): 3577–3599.

Woodward, J. (2015b). "Interventionism and Causal Exclusion." *Philosophy and Phenomenological Research* 91 (2): 303–347.

Woodward, J. (2016). "The Problem of Variable Choice." *Synthese* 193 (4): 1047–1072.

Woodward, J. (2021a). "Explanatory Autonomy: The Role of Proportionality, Stability, and Conditional Irrelevance." *Synthese* 198 (1): 237–265.

Woodward, J. (2021b). *Causation with a Human Face*. Oxford University Press.

Wunsch, G., M. Mouchart, and F. Russo (2014). "Functions and Mechanisms in Structural-Modelling Explanations." *Journal for General Philosophy of Science* 45 (1): 187–208.

Yablo, S. (2002). "De Facto Dependence." *Journal of Philosophy* 99 (3): 130–148.

Yablo, S. (2004). "Advertisement for a Sketch of an Outline of a Proto-Theory of Causation." In *Causation and Counterfactuals*, edited by J. Collins, N. Hall, and L. Paul, 119–137. MIT Press.

Yagan, D. (2019). "Employment Hysteresis from the Great Recession." *Journal of Political Economy* 127 (5): 2505–2558.

Yee, A. S. (1996). "The Causal Effects of Ideas on Policies." *International Organization* 50 (1): 69–108.

Yule, G. U. (1926). "Why Do We Sometimes Get Nonsense-Correlations between Time-Series? A Study in Sampling and the Nature of Time-Series." *Journal of the Royal Statistical Society* 89 (1): 1–63.

Zhang, J., and P. Spirtes (2008). "Detection of Unfaithfulness and Robust Causal Inference." *Minds and Machines* 18 (2): 239–271.

Zhang, J., and P. Spirtes (2014). "Choice of Units and the Causal Markov Condition." In *Scientific Explanation and Methodology of Science: Selected Papers from the International Conference on SEMS 2012*, edited by G. Guo and C. Liu, 240–251. World Scientific.

Zhang, C., K. Mohan, and J. Pearl (2022). "Causal Inference with Non-IID Data Using Linear Graphical Models." *Advances in Neural Information Processing Systems* 35: 13214–13225.

Zhao, S. (2024). "A Causal Theory of Confirmation for Bayesians." *Synthese* 204 (6): Article 167.